REPORT RESUMES

ED 003 346 24

SEQUENCE CHARACTERISTICS OF TEXT MATERIALS AND TRANSFER OF LEARNING, PART 1--EXPERIMENTS IN DISCOVERY LEARNING.

BY- DELLA-PIANA, GABRIEL M. AND OTHERS
UTAH UNIV., SALT LAKE CITY, BUREAU OF EDUC. RES.
REPORT NUMBER CRP-2277-PT-1 PUB DATE DEC 65
EDRS PRICE MF-$0.36 HC-$8.12 203P.

DESCRIPTORS- *LEARNING PROCESSES, *COMPARATIVE ANALYSIS, *MATHEMATICS INSTRUCTION, *INSTRUCTIONAL MATERIALS, TEACHING TECHNIQUES, *PROGRAMED TEXTS, SEQUENTIAL PROGRAMS, GRADE 5, GRADE 6, GRADE 9, ALGEBRA, SALT LAKE CITY, UTAH

THE TRANSFER EFFECTS OF DISCOVERY AND EXPOSITORY INSTRUCTIONAL TECHNIQUES FOR SEQUENCING OF INSTRUCTION WERE THE PURPOSE OF THIS TWO-PART STUDY. THE FIRST STUDY COMPARED THE TWO PROCEDURES IN A PROGRAMED UNIT ON SUMMING NUMBER SERIES. SAMPLES FOR THIS PART OF THE STUDY CONSISTED OF 96 NINTH-GRADE ALGEBRA STUDENTS, WHO WERE ASSIGNED TO EITHER OF THE TWO TREATMENTS. THE SECOND STUDY COMPARED THE PROCEDURES WITH TEACHER PRESENTATION ON SELECTED MATHEMATICAL CONCEPTS. SAMPLES CONSISTED OF 538 FIFTH- AND SIXTH-GRADE STUDENTS WHO RECEIVED INSTRUCTIONS BY THE TWO METHODS AND NORMAL INSTRUCTION. DATA WERE OBTAINED FOR TIME AND ERROR SCORES ON THE PROGRAM, AND RETENTION AND TRANSFER MEASURES. ANALYSES INVOLVED SIMPLE COMPARISONS OF MEANS AND VARIANCE. THE RESULTS INDICATED THAT THE GUIDED DISCOVERY SEQUENCING AND METHOD WAS SUPERIOR IN TRANSFER EFFECTS AND RETENTION OF CONCEPTS TO EXPOSITION. FURTHER RESEARCH SUGGESTIONS ARE DIRECTED TOWARD TREATMENT-APTITUDE INTERACTIONS AND USING APPROACHES TO ELICIT PROBLEM-SOLVING PROCESSES. THIS REPORT DESCRIBED THE DETAILED PROCEDURES. SAMPLE COPIES OF THE MATERIALS USED WERE ASSIGNED ACCESSION NUMBER ED 003 347. (RS)

ED003346

CR 2277
PT I

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

SEQUENCE CHARACTERISTICS OF TEXT MATERIALS AND TRANSFER OF LEARNING

PART I: EXPERIMENTS IN DISCOVERY LEARNING

GABRIEL M. DELLA-PIANA, Project Director

AUTHORS:

GABRIEL M. DELLA-PIANA, GARTH M. ELDREDGE, BLAINE R. WORTHEN

BUREAU OF EDUCATIONAL RESEARCH
UNIVERSITY OF UTAH • SALT LAKE CITY, UTAH 84112

DECEMBER 1965

The research reported herein is Part I of the Final Report of Project 2277 supported under the Cooperative Research Program of the Office of Education, U.S. Department of Health, Education and Welfare

ACKNOWLEDGMENTS

The effective conduct of this study would have been impossible without the support of the following:

The U. S. Department of Health, Education, and Welfare, Office of Education for financial support through the Cooperative Research Program, Project No. 2277.

Pupils and personnel of the Granite School District and the Salt Lake City School District in Salt Lake City, Utah.

The Stewart School Faculty, the University of Utah Laboratory School on Campus for providing a compatible climate for pilot work.

Miss Marilyn Latham, Mrs. Prudence Redford and Mrs. Marjean Ballard for participation in data gathering and analysis.

Stanley M. Jencks for unstinting and invaluable consultant help in the writing of the materials for the research reported in Chapter V.

Mrs. Carolyn Bennion for her excellent secretarial services.

TABLE OF CONTENTS

CHAPTER PAGE

I. INTRODUCTION, ABSTRACTS OF CHAPTERS AND IMPLICATIONS 1.00

II. LITERATURE REVIEW 2.00

III. TASK ANALYSIS 3.00

IV. DISCOVERY VERSUS EXPOSITORY SEQUENCING IN A PROGRAMED UNIT ON SUMMING NUMBER SERIES 4.00

V. DISCOVERY VERSUS EXPOSITORY SEQUENCING IN SIX WEEKS OF CLASSROOM INSTRUCTION IN MATHEMATICS 5.00

CHAPTER I

INTRODUCTION, ABSTRACTS OF CHAPTERS, AND IMPLICATIONS

G. M. Della-Piana

The studies reported here were designed to yield information on the relative transfer effects of different procedures for sequencing of instruction. The procedures investigated have been loosely referred to as "guided discovery versus direct exposition of a generalization" although those terms are far too weighty or at least too ambiguous for the experimental arrangements we have contrived.

We have been careful to design the studies in such a way as not to be subject to the typical criticisms of "methods" studies (Travers, 1962, p. 540) which have also been levelled at research in the field of programed instruction (Stolurow, 1960, p. 2). Our procedures are described in detail in this report and instructional materials used in the studies are presented in a separate volume of appendices.[1]

This chapter presents brief abstracts of the chapters in this volume and a section on implications for further research.

Abstracts of Chapters

Chapter II. Summary of Research Literature.--

Seventy-two studies are reviewed, for the most part, in a chronological order. Other analytic comments on the literature are included in

---

[1] G. M. Della-Piana, Blaine R. Worthen and Garth M. Eldredge, Sequence Characteristics of Text Materials and Transfer of Learning, Part II. Appendices: Tasks and Tests. Salt Lake City: Bureau of Educational Research, University of Utah, 1965.

Chapter I and Chapter III. The differing operations for defining sequence characteristics of text materials is so staggering as to make it impossible to summarize trends in the literature. This chapter, however, provides an excellent background for the critical material in the following chapter.

Chapter III. Task Analysis.--

Growing interest and activity in task analysis is cited. The difficulty in understanding the significance of even a carefully designed piece of research on sequence characteristics of instructional materials is documented by analyzing the materials used in the Gagné-Brown study. The lack of interpretability of the study is discussed. The Guided Discovery treatment is shown to have advantages in association value and fading of prompts which were uncontrolled relevant nonexperimental variables. The chapter concludes with a description of attempts to control relevant nonexperimental variables while varying the crucial experimetal variables in developing materials for the studies reported in Chapters IV and V. Samples of the materials are included. The complete set of materials is published in Part II of this final report published as a separate volume.

Chapter IV. Discovery Versus Expository Sequencing in a Programed Unit on Summing Number Series.--

A modified replication of the Gagné-Brown (1961) study is reported briefly. Though some changes in procedure were introduced the results supported Gagné-Brown's findings directionally on total problems and significantly on problem one of four transfer problems. That is, students learning to sum number series in a guided discovery (GD) program performed better on transfer problems than those given a rule and example (RE) or direct exposition treatment. The major purpose of the study was to provide data for revision of the programs.

Using revised forms of the GD and RE programs for summing number series, as described in Chapter III, Eldredge conducted a study which attempted to compare these treatments when the major experimental variable was whether one begins with a copying frame and fades away to producing the rule for summing and the sum of number series or whether one begins by producing sums and trying to formulate a rule and then later being given a copying frame. The major findings were that there were no sex differences, no ability by method interaction (although all subjects were relatively high in ability) no immediate transfer test differences, but significant delayed (4-week) transfer test differences favoring the GD treatment. The delayed transfer effects disappeared when initial transfer test was used as a covariate suggesting that it was initial learning that produced the differences. Subjects allowed to explain orally their problem solving processes after the immediate transfer test showed greater gains on delayed transfer than those not given this opportunity.

An exploration study of problem solving processes is reported in which for the part of the sample studied no treatment differences were found on problem solving processes. Since data on students selected for this data gathering had not yet been analyzed for the immediate or delayed transfer test it was not known how they performed on it. It so happened that for this subgroup there were no treatment differences on immediate or delayed transfer,thus leaving open the question as to whether problem solving differences would have appeared had we tested subjects who contributed to the significant differences in delayed transfer in the total sample.

Chapter V. Discovery Versus Expository Sequencing in Six Weeks of Classroom Instruction in Mathematics.--

In an attempt to measure the same general variables as the previous studies over a longer treatment period, a six week mathematics unit was developed for sixth graders. Instructional materials from three major curriculum projects were adapted to prepare units on real number addition and multiplication, the distributive principle of multiplication over addition, exponential notation and multiplication and division of numbers expressed in exponential form. Although material was prepared in written form, the treatment made was teacher presentation, each teacher handling one discovery and one expository class. Analysis of teacher behavior by observer rating and pupil rating showed high fidelity to treatment models. The model of discovery teaching included dimensions, such as: 1. Teacher avoiding attitude that he knows all answers; 2. Delay of verbalization of the concept until a minimum of 85 percent of the pupils made a discovery; and 3. Answering questions by referring the pupil back to the material. There was careful control of time on task, amount (not sequence) of teacher verbalization, "Hawthorne Effect," and teacher fidelity to the treatment model. The major findings were that the discovery method proved superior to the expository method in retention of mathematical concepts, transfer of heuristics to new task situations and attitude toward arithmetic. The expository treatment was superior on initial learning o mathematical concepts. Thus, this study strongly supported the value of discovery over expository teaching and demonstrated the feasibility of training teachers to assume two teaching styles with excellent fidelity to the model.

Implications

The studies reported above present considerable support for transfer advantages to instruction sequenced to call for searching for principles before being presented with the principles. The characteristics of the material producing these results is clearly documented. Though the trend is fairly clear, the exceptions to this generalization provide clues for further research.

There has been a continuing concern about the merits of control group, methods studies where the two treatments are "unspecified amalgums of several principles" (Stolurow, 1960) or even specified amalgums of several principles (Cronbach, 1965). When you have complex treatments, you don't know what characteristics of the treatment produce the observed results. This observation has led Cronbach to argue that the best methods comparison is one in which there is optimization of each method within a fixed time so that the best example of each method is compared even if the content itself is not held constant. In other words, if it is almost impossible to know what characteristics of the method is producing the results, why not eliminate experimentation and resort simply to evaluation studies of complex treatments so you can say what it is that the method does do? I should like to extend that argument further and say that the method (though complex) should be specified so that other curriculum developers or researchers can use the specifications to produce another set of materials with that complex set of characteristics. Then, if they can replicate the results obtained in previous studies, you will have demonstrated generalizability of the "complex method" even though you may forever be in the dark as to what specific characteristics of the method produced the results.

It has been mentioned earlier in the report that the control group type of study reported here "destroys the treatment." I would argue that such destruction is valuable in isolating treatment parameters that are worth further attention. But it is probably true that the general effect of such studies is often to make it impossible to find some crucial variables by ruling them out of the design. Suppose we were to pursue the line of research begun in this project by attempting to optimize a treatment. How would we proceed differently from this point on? The following suggestions seem to me to provide fruitful directions for further research along these lines.

1. Use currently available data to generate promising hypotheses concerning treatment-aptitude interactions so that a treatment can be optimized for a given aptitude pattern. The off-quadrant analysis of data recently described in some detail provides a useful technique for making use of current data to generate empirically basedhypotheses (Marks, 1964). For example, in one of our pilot studies math aptitude was significantly positively related to performance on total problems in the discovery treatment (r = .42) and significantly negatively related to total problems for the expository treatment (r = -.33). Data such as this could be analyzed impressionistically to see whether any other data on the subjects could provide clues to explain this phenomenon. Hypotheses thus generated could then be subjected to experimental test.

2. Observe the behavior of interest in "natural" situations. It is clear that the controls built into our studies made it virtually impossible to study "adventurous thinking" (Bartlett, 1958) or problem solving in which" the problem exists but remains to be identified or discovered, and no standard method for solving it is known to the problem solver or to the others (Libby, 1964) (Mackworth, 1965). Our data is what Barker

(1965) calls transducer generated. That is, we have imposed a structure on the observational situation. Following procedures for observing behavior in "natural" situations (Barker, 1965) (Weir, 1962) or simulations that closely approximate such situations may uncover variables or relationships we had not clearly identified before. These variables and relationships could then be the basis for formal experimentation. Or a theory of information processing derived from protocols thus obtained could be programed for computer simulation and tested by comparing the behavior of the computer with the behavior of subjects when both are given the same problem (Newell and Simon, 1961).

3. Specify the structure and sequence characteristics of a complex treatment. Develop the treatment until it produces the behavior desired at an optimal level. Develop other materials following your specifications until you have demonstrated the generalizability of your specifications by being able to produce with new material the optimal level of student performance by following your procedures. Finally, communicate these procedures to others and if they too can produce the behavior you have produced you will have demonstrated the generality of your procedures over subject matters and curriculum writers.

4. Finally, the procedures for eliciting and categorizing problem solving processes are developing rapidly as is evidenced by the increasing number of studies of this type. (Bloom and Broder, 1950; Bloom, 1953; Buswell, 1956; Connolly and Wantman, 1964; Gallagher, 1961; Glaser and others, 1952; Kersh, 1965; McGuire and Babbot, 1965; Saupe, 1954; Shulman, 1965; and Siegel, 1963). The data-gathering innovations represented by these studies warrant considerable use and give promise of uncovering problem solving process categories and relationships not now clearly specified.

The implications drawn above have all been in the direction of further research and materials development procedures. An immediate practical implication of the research is that the transfer advantages of a "discovery" treatment can be obtained without any time-cost disadvantage. This is an implication which contradicts previous speculations in the literature.

References

Barker, R. G., "Explorations in ecological psychology," Amer. Psychol., 1965, 1, 1-14.

Bartlett, Sir Frederic, Thinking: An Experimental and Social Study. Basic Books, Inc., Publishers, New York, 1958.

Bloom, B. S. and Broder, Lois J., "Problem solving processes of college students," University of Chicago Press, 1950. (Supplementary Educational Monographs, No. 73.)

Bloom, B. S., "Thought processes in lectures and discussions," J. gen. Ed., 1953, 7, 160-169.

Buswell, G. T., Patterns of Thinking in Solving Problems. University of California Press, Berkeley and Los Angeles, 1956.

Connolly, J. A. and Wantman, M. J., "An exploration of oral reasoning processes in responding to objective test items, J. educ. Msmt., 1964, 1, 59-64.

Cronbach, L. J., The Logic of Experiments on Discovery. Paper presented at a conference in New York City, January 28-29, 1965. Supported by the U. S. O. E. Cooperative Research Program.

Della-Piana, G. M., "Searching orientation and concept learning," J. educ. Psychol., 1957, 48, 245-253.

Gagné, R. and Brown, L. T., "Some factors in the programing of conceptual learning," J. exper. Psychol., 1961, 62, 313-321.

Gallagher, J. J. and others. A System for Classifying Thought Processes in the Context of Classroom Verbal Interaction. (Ditto) Inst. for Research on Exceptional Children, University of Illinois, July, 1961.

Glaser, R., Damrin, Dora E., and Gardner, F. M., The Tab-Item: A Technique for Measurement of Proficiency in Diagnostic Problem Solving Tasks. Urbana: University of Illinois, Bureau of Research and Service, College of Education, 1952.

Heidbreder, E., "The attainment of concepts: III. The processes," J. of Psychol., 1947, 24, 93-138.

Kersh, B. Y., "The motivating effect of learning by directed discovery," J. educ. Psychol., 1962, 53, 65-71.

Kersh, B. Y., Directed Discovery Vs. Programed Instruction: A Test of a Theoretical Position Involving Education Technology. Paper read at American Educational Research Association Meetings, Chicago, February 12, 1965.

Libby, W. L., "Tools for discovery of problems and their solutions," Quoted in J. W. Getzels' "Creative thinking, problem-solving, and instruction," in E. R. Hilgard (ed.) Theories of Learning and Instruction. 63rd Yearbook, NSSE, Part I, 1964, University of Chicago Press, Chicago, Illinois.

MacKworth, N. H., "Originality," Amer. Psychol., 1965, 1, 51-66.

Marks, M. R., "How to build better theories, tests, and therapies: The off-quadrant approach," Amer. Psychol., 1964, 19, 793-798.

McGuire, Christine H. and Babbott, D., Simulation Technique in the Measurement of Problem-Solving skills: A Technological or Conceptual Breakthrough. Paper presented at the joint AERA-NCME meeting, Chicago, February, 1965.

Newell, A. and Simon, H. A., "Computer simulation of human thinking," Science, 1961, 134, No. 3495, 2011-2017.

Reitman, W. R., "Information processing models in psychology," Science, 1964, 144, 1192-1198.

Saupe, J. L., Trouble Shooting Electronic Equipment. Urbana, University of Illinois, Bureau of Educational Research, College of Education, 1954.

Shulman, L. S., Seeking Styles and Individual Differences in Patterns of Inquiry. Paper read at meetings of American Educational Research Association, Chicago, February 12, 1965.

Siegel, L. and others. "Students thought during class: A criterion for educational research," J. educ. Psychol., 1963, 54, 45-51.

Stolurow, L. M., "Teaching machines and special education," Educ. and Psychol. Msmt., 1960, 20, 429-448.

Travers, R. M. W., "A study of the relationship of psychological research to educational practice," in R. Glaser (ed.) Training Research and Education. Pennsylvania: University of Pittsburgh Press, 1962.

Weir, Ruth H., Language in the Crib. The Hague: Mouton, 1962.

CHAPTER II

SUMMARY OF RESEARCH LITERATURE

Research on discovery learning has been in process for several decades. Proponents of the discovery approach have probably been influenced in part by research and opinion growing out of the Progressive Education Movement of the twenties. A number of modern proponents of the discovery method recognize this heritage, but see differences between the discovery methods of today and the inductive methods of the twenties.

A review of the literature on learning by "discovery" quickly shows that there is no unanimity among researchers in their use of the term. There are, of course, rudimentary commonalities underlying each of the various conceptions of discovery, and this perhaps permits us to use the term rather loosly when referring to these basic similarities.

Even within the framework of agreement concerning what elements are generally characteristic of discovery learning, there is still another divisive agent which quickly becomes apparent when one attempts to review the literature. This stems from the fact that there are many variables which are operating and being operated on simultaneously under any conditions of discovery. A review of the research literature shows that many of these variables have been explored to some degree, while others have received relatively little attention. This wide divergence of the variables controlled in various studies has further compounded the difficulty of drawing logical inferences from the research findings. The experiments that have been conducted have resulted in seemingly inconsistent results. This apparent inconsistency seems to reflect not only differing conceptions among experimenters of what learning by discovery is, but also an investigation of widely differing facets of the discovery process. In addition, many of the studies have failed to specify the sequence characteristics of the learning tasks, thus making it difficult to generalize from what research has been done. These are merely indicative of many problems which make the research extremely difficult to interpret.

The lack of clarity in the definition of terms and the resulting misleading labeling of various methods and experimental procedures, as well as the many possible combinations of variables used in these studies, prevents any uniform organization of the literature review. The literature will, therefore, be reviewed chronologically for the most part in this chapter. Chapters I and III include a more analytic and critical look at portions of related research.

In one of the earlier studies related to discovery learning, Ewert and Lambert (1932) undertook the comparison of four teaching methods which varied in the amount of teacher guidance. The learning task was a series of problems which required the subjects to move different sized disks about among three circles until the arrangement met specified criteria. A principle could be employed in the solution of the problem. Method 1 gave the subjects the objectives of the problems and the procedural rules. Method 2, in addition to stating the objectives and procedural rules, asked the subjects to find the one general principle which was appropriate to all the problems. Method 3 gave the objectives, the rules, and the general principle; while Method 4 was identical to Method 3, except for the addition of a demonstration of the principle. The criteria of the number of moves and the time to reach the solution were used. The resultant ranking of the methods, in order, was 3, 4, 2, and 1. The authors concluded that the more complete the verbal instructions were made, the less manipulation was necessary to solve the problem. They also concluded that the reversal of order between Method 3 and 4 was due to an inhibiting effect of the demonstration on the subjects taught by Method 4.

In one of the earlier studies related to discovery learning, McConnell (1934) taught second grade subjects one-hundred addition facts and one-hundred subtraction facts by two methods: authoritative and discovery. The authoritative group was simply told to memorize the facts with no attempt to give the subjects understanding. The discovery group was told to discover the principle involved in the learning task. The training lasted for seven months. McConnell found that subjects taught by the discovery method were able to transfer the learning to a test of untaught number combinations better than those taught by an authoritative method. Those taught by the authoritative method were superior to the discovery group on speeded retention tests, however. McConnell suggested that the discovery learning teaches a student how to think, and is therefore a more general and hence more valuable type of learning.

Thiele (1938) used a research design almost identical to McConnell's (1934) in comparing a "generalization" method with a "memory-drill" method of learning 100 addition facts. The memory-drill approach emphasized mechanical repetition and rote memorization of the facts, whereas the generalization method included derivation and verbalization of the generalization. Thiele found the generalization method clearly superior.

Katona (1940) performed two series of experiments in which he contrasted methods of "memorizing" with methods of "understanding." In the first series, 100 subjects were required to learn a series of card tricks based upon certain principles. He found that subjects who had discovered the principles without any assistance scored highest on retention and transfer tests. Subjects who had watched the solutions and had been told the principles showed the next highest gains, while the "memorizing" group scored lowest among his experimental groups. A control group receiving no instruction showed least gains.

These findings led Katona to his second series of experiments in which the formation of geometrical patterns from matchsticks comprised the learning task. Three methods were compared. A "memorizing" method consisted of repetitious demonstration and explanation of the solutions to the task. The second method, a "help" or understanding method, allowed the subjects to discover the solutions with the aid of illustrative hints and demonstrations to explain the sequence of steps in the solution. Later, a third method was added in which the principle was stated, the solution shown, and the principle restated. Katona found the "help" method significantly better than the "memorizing" method and slightly better than the third method on transfer tests given immediately after treatment and again after a four-week interval.

The hypothesis that failures in solving a problem promote inductive reasoning more than immediate and uniform successeswas tested. Forty-two subjects, divided into two groups, were presented with a series of cards, each requiring the pressing of one button on a three-button apparatus. The problem consisted in the subject's discovering which of six cues to follow in order to make a bell ring. Results show that subjects in the restricted hypothesis group profited by their mistakes and surpassed the performance of the unrestricted hypothesis group by a statistically significant margin.

Hendrix (1947) compared three methods of presenting a mathematical generalization to three small groups of high school and college students.

In Method 1, the principle was stated, clarifying illustrations were given, the principle was restated, and then subjects were asked to apply it to several examples. In Method 2 the subjects were given a series of problems which led to self-discovery of the principle. As soon as each subject recognized the principle, he was dismissed. Method 3 was identical to Method 2, except that each subject was also required to give a correct verbalization of the principle immediately following his discovery of it. Two weeks after learning the generalization, the three groups were given a transfer test. Hendrix found that the methods ranked, in order of effectiveness, 2, 3, and 1. Although Hendrix's studies were merely pilot experiments which failed to reach significance beyond the .12 level, it is interesting to note that the trend in all three studies indicated superiority for the nonverbalized discovery technique.

Schwartz (1948) conducted a study of the importance of verbalization in concept formation. The learning task consisted of sorting blocks and cards according to several complex sorting principles. The subjects were fifty adults of generally high intelligence. Subjects were tested on a series of sorting experiments which differed in the necessity of verbalizing the sorting principle. Schwartz found that: (a) most subjects who learned and successfully transferred a sorting concept were still unable to verbalize the principle they were using; (b) these same subjects showed, after the unsuccessful attempt at verbalization, a marked decrease in ability to transfer the principle to a new situation; and (c) the few subjects who successfully verbalized the principle were always successful on a subsequent transfer test. Schwartz concluded that concepts can be formed and used on a subverbal basis, that unsuccessful attempts at verbalization may destroy the concept, and that successful attempts at verbalization facilities future use of the concept. His finding have been interpreted as lending strong support to the conclusions of Hendrix (1947).

Two long-term curriculum studies (Swenson, 1949; Anderson, 1949) were similar to the studies of McConnell (1934) and Thiele (1938) both in terms of the length of the experimental period and in terms of the research design. Swenson found that second-grade pupils who had been taught by a generalization method performed significantly better on retention and transfer tests than pupils taught by a drill method. Anderson found no differences in recall between the two treatments but found his "meaningful" method to be superior over the "drill method" in terms of transfer. His experiment was conducted at the fourth-grade level.

Stacey (1949) compared the relative effectiveness of five methods in which different amounts of guidance were employed. The task required subjects to identify the one word in each of several sets of five that did not "belong" to that particular set. Two methods were considered to be discovery methods; one a directed discovery method and the other an independent discovery method. The other three methods represented varying degrees of "authoritative" techniques. In all three of these methods, the subjects were given the correct responses to the problems in some greater or lesser degree. Stacey concluded that the two discovery methods contributed "better" learning and that errors made under the discovery conditions tended to be less detrimental than errors made under the other three conditions. However, no significant differences on the pre- to post-test gains were found between groups taught by these two basic methods.

In the first of two studies, Craig (1953) utilized a learning task very similar to that used by Stacey (1949). Unlike Stacey's study, however, none of the methods employed by Craig provided responses to specific terms, although varying degrees of guidance existed in the four methods.

The task involved the learner's discovering which one of five words did not belong when the other four were grouped according to some principle. His items were designed for use with college graduates and were at a high level of difficulty.

The four treatments consisted of: (a) no guidance; (b) guidance wherein each group of items based on an organizational principle was separated from others by spacing; (c) guidance consisting of the grouping of items as in "b" plus notice that items were grouped according to a principle; (d) maximum guidance wherein subjects were given the same guidance as that in "c" plus a statement of the principle involved.

Subjects included in his study were recent college graduates (50 in each group) who were equated on ability levels based on a series of pre-tests. The pre-tests were given again as post-tests to assess the amount of learning which took place. Results of the study supported the notion that increased guidance aids retention and transfer. Subjects who were given maximum guidance performed significantly better than all others. Conclusions were: (1) the amount of transfer increases as more and more cues are provided to aid discovery; and (2) the transfer effect of guidance becomes greater as the difficulty of the situations to which learning is transferred increases.

Craig (1956) followed his first study with a second wherein two groups of college students were given the same tasks as those involved in the earlier study. His "independent discovery" group was given the direction, "Different principles are used for different groups," and subjects were presumably to discover the principle of grouping. The "directed discovery" group was given the same directions as the independent discovery group except that after the last statement was read, the subjects were told, "The principle applying to each group of four items is stated just above the group." A short statement concerning the principle was stated just above the group of words.

Subjects were given pre-tests and three days later were given a learning task (1st day) which consisted of learning two illustrative items for each of five groups of words each with a different organizational basis (set A). Two days later (3rd day) subjects were required to discover two additional correct responses to each of the five organizationally based groups in A and two days later (5th day) were required to discover four illustrative items to each of the groups in A. On the 15th, 17th and 19th days of the study, subjects repeated the procedure in A for a new set of tasks based on five more organizational principles (set B) and, on the 29th, 31st and 33rd days repeated the experiment with still a third set of tasks based on five additional organizational principles (set C). In all, subjects learned tasks based on 15 different organizational principles. Post-tests were administered three days after completing set A.

The directed discovery group learned significantly more relations than the independent discovery group on each of the three trials. The groups did not differ in the per cent of learned relations retained after 3 and 17 days, but after 31 days the directed discovery group retained significantly more relationships than the independent discovery group. Post-tests of transfer indicated that the groups did not differ significantly in the amount of transfer to unlearned relationships, although the direction favored the directed discovery group.

A study by Judson, Cofer and Gelfand (1956) was designed to determine the influence of prior reinforcement of relevant responses on problem solving of college students. Three problems were used: two string problem, the Maier hat rack problem and a verbal problem set up in the light of free associations obtained from the subjects. Prior to the problem solutions, several lists of words were learned by the subjects in the setting of a retroactive inhibition experiment. In one of the word lists, a series of words relevant to problem solving occurred together for one group. Following this learning, the problem was worked on, and then a recall for the word lists was obtained. It was found that the relevant patterns of verbal associations, set up by learning in the first stage of the experiment, were associated with the frequencies of certain types of solution in the two string and the hat rack problems. In the verbal problem it was found that problem solution requiring choice of members of a chain of free associations was affected by prior reinforcement of one member of the chain.

A series of comparative studies in mathematics (Sobel, 1954; Norman, 1955; Fullerton, 1955; and Pincus, 1956) all compared "traditional" with "discovery" methods of teaching.

Sobel (1954) studied the effects of two methods of teaching ninth-grade algebra to pupils with high IQ. He concluded that the discovery method yielded better results than the traditional didactic approach.

Norman (1955) found her "developmental" approach (which allowed for discovery) to be superior to the conventional expository approach in terms of retention and transfer.

Fullerton (1955) used the teaching of multiplication facts in the third grade to compare an inductive method with a deductive method. Fullerton found the inductive method superior both in terms of initial learning and in terms of subsequent retention and transfer.

Pincus (1956) compared a rote-memorization method with an understanding method of learning subtraction and addition facts. This study produced no significant differences in favor of either method.

Two experiments were conducted by Ackerman (1956) to determine the differences in problem-solving behavior between subjects who have been trained in alternative methods of solutions and subjects who had not received such training. In experiment I, 39 sixth grade students were divided into 3 groups of 13 subjects each. The experimental group (Group A) was taught two different ways of solving each problem. The two control groups were taught only one way of solving the problems. The three groups were matched on intelligence and subject matter achievement. Problems selected included a water-jar problem, a maze problem, and a paper and pencil puzzle problem. There were no significant differences in achievement among the three groups. A similar experiment was conducted again with 48 sixth graders. In this experiment, a group of subjects who were taught that there is more than one way to solve a problem showed more variable solutions to a similar problem than a group that was taught only one solution to the original problem. Finally, male subjects preserved and transfered their ability longer than females.

Kittell (1957) reported an experiment in which three amounts of guidance were used in presenting the principles to sixth-grade pupils. The learning task was similar to that used by Craig (1956). The minimum-direction group was given examples of the principles involved in the task and was told only that there was a principle involved. In the intermediate-direction treatment, in addition to the examples, the principle was stated for each group of examples. The maximum-direction group received all the intermediate guidance plus verbal statements of the correct answers to each question. Kittell found that the intermediate-direction method was superior to the maximum and minimum methods in terms of retention and transfer. The evidence from this study was interpreted as supporting Craig's (1953, 1956) conclusions that presentation of principles, without giving the correct answers, proved superior to self-discovery of the principles.

Corman (1957) studied the effects of varying the kinds and amounts of information given to subjects in a learning task. He used Katona's match stick problems wherein subjects were required to form new geometrical designs from old ones by appropriate manipulation of certain numbers of matches.

Subjects were given varying amounts of information (a) as to the rule involved and (b) as to the method involved. Information about the rule consisted of (a) no information, (b) some information (subjects were given a statement concerning the number of matches involved and told that a rule existed and that they were to look for the rule) and (c) much information (subjects were told that a rule existed and were told the rule). Varying amounts of information concerning the method were also given: (a) no information, (b) some information (subjects were directed to shade in certain squares and their attention was drawn to the number of matches remaining in the unshaded square) and (c) much information (subjects were told which matches to move where and were told to practice the solution).

Combinations of the information about the rule and method produced nine possible instructional variations as follows: (a) none on rule, none on method; (b) none on rule, some on method; (c) none on rule, much on method; (d) some on rule, none on method; (e) some on rule, some on method; (f) some on rule, much on method; (g) much on rule, none on method; (h) much on rule, some on method; (i) much on rule, much on method. Each variation was considered as a separate treatment.

Subjects were given an introduction, a trial problem, an explanation of restrictions, and an instructional task. They were then given equal amounts of time to work a series of test problems after which they were asked to write a statement about the rule. Success was appraised in the number of instructional problems solved, in transfer to simple and complex problems, and in verbalizing the principle.

For subjects above average in mental ability success in solving the instructional problems increased as the amount of information about the method increased. For low ability subjects some appropriate guidance on the method was superior to no guidance. Information about the rule did not affect the results differentially.

For simple transfer problems, the high ability subjects profited most from the combination of much information about the rule and some information about the method. On complex transfer tasks, high ability subjects receiving only information as to the method or only an indirect clue as to the rule did less well than subjects receiving other combinations. On these complex tasks, low ability subjects who were given only guidance as to the method (much or some information) were superior.

Success in writing the rule increased as the amount of information about the rule increased.

The study was interpreted to indicate that (a) information used in guidance must be appropriate to the subject, (b) some appropriate guidance is beneficial, but failure to provide such information will delay rather than prevent solution, (c) the effectiveness of guidance does not depend solely on the amount of information imparted, (d) more explicit information will be more helpful with high ability subjects and (e) less explicit information (i.e., allowing subjects to manipulate the materials rather than giving specific direction) may be just as effective as directive guidance for the less able subjects.

Forgus and Schwartz (1957) performed an experiment to test the validity of Katona's findings and, if possible, to improve the "understanding method." Three groups of 13 subjects were required to learn a new 26-symbol alphabet. One group (observer-O) was given written instructions explaining the principle underlying the construction of the new alphabet. The second group (participant-P) was told there was a principle and asked to describe what it was. The third group (memorization-M) was given the alphabet, randomly ordered, and told to learn it letter by letter. After a week, tests were given for recall, transfer, and problem solving. Groups O and P were tremendously superior to group M on all three tests ($p < .001$). No significant differences were found between methods O and P on any of the three measures.

This study by Forgus and Schwartz has been interpreted (Kersh and Wittrock, 1962) as involving a discovery group (P), a guided discovery group (O), and a rote-learning group (M). More careful examination of the learning task seems clearly to indicate that Method O was not given an opportunity to discover the principle; rather, group O received a meaningful exposition of the principle. Therefore, this study should not be quoted in defense of either discovery or expository learning, as it merely compares a rote method with two understanding methods (one discovery and one expository), both of which were significantly superior to the rote method.

Della-Piana (1957) investigated the learning effects of two methods of feedback, for college age subjects. Subjects were either told the correct answer when an incorrect response was made (dependency treatment) or were told to keep trying until the right answer was discovered (searching treatment). The learning task was an inductive concept identification task wherein subjects were to learn the names (nonsense syllables) of concept drawings presented singly but successively in series after series of drawings. "Searching" treatment subjects recalled and recognized a significantly greater number of meanings than the "dependency" treatment subjects. The "searching" treatment appeared to enhance the learning of meanings.

Hasslerud and Myers (1958) conducted a study using coded words in a sentence. A total of twenty principles were involved in the coding experience. Two methods of coding were used: given (G) wherein the coding principle was stated and examples were decoded, and derived (D) wherein the subject was required to discover the principle from examples which were given. Each subject learned ten principles by each method, and one week later a transfer test in decoding was administered. Initial learning favored the given group but after one week, there was no difference between groups. However, the initial scores earned in the learning process and the transfer scores were added algebraicly to give a gain score for the "derived" principles and a net loss for the "given" principles. The difference between methods using this score was significant. The authors concluded that independently derived principles were more transferable than those where the principle was given. The latter interpretation has been subjected to criticism, however, on the basis that such an inference cannot be drawn from the data. As Ausubel says:

> . . . this in my opinion is equivalent to saying that of two matched race horses trained by methods A and B respectively, who are tied at the end of the criterion race, the horse trained by method B is really superior because at the halfway mark he was one lap behind the horse trained by method A, but nevertheless, caught up to him by the end of the race. (Ausubel, 1963, p. 170)

Kersh (1958) conducted a study with college students to test the notion that learning by independent discovery is superior to learning with direction because learning is more meaningful in the former case. Meaningful learning is used in the cognitive sense of organization and understanding. Each subject was required to learn two rules of addition: (a) the odd numbers rule (the sum of any series of consecutive odd numbers beginning with one is equal to the square of the number of figures in the series); and (b) the constant-difference rule (the sum of any series of numbers in which the difference between the numbers is constant is equal to one-half the product of the number of figures in the series and the sum of the first and last numbers). Subjects were divided into three groups: (a) no help; (b) direct reference (some direction in the form of perceptual aids); and (c) rule given (a verbal statement of the rule was given to the subject). Problems were of two types: (a) numerical and (b) geometrical giving a total of six possible treatment combinations. On a test of learning immediately following the learning program and on a retest after four weeks, the no help treatment groups was superior to the other treatment groups. An analysis of subject reports on their activities in the four weeks indicated that subjects receiving no help practiced the problems more outside class than the other subjects. Kersh reasons that as a result the learning became more meaningful and hence, more transferable.

Kersh (1962) investigated this phenomenon further in a study with a group of high school subjects. The same rules of addition as were taught in the earlier study were used again. Ninety subjects participated in the study and were divided into three groups: Group A was given individual guidance in the discovery of the explanation involved (guided discovery); Group B was taught the explanation by a programed textbook (directed discovery); and Group C was given no further direction or explanation of the rules. The rules were merely presented for memorization (rote-learning).

A questionnaire and test of recall were administered after three days, two weeks, and six weeks. The tests favored the rote-learning and guided discovery groups. The questionnaire indicated that the guided discovery treatment group practiced the rules during the time interval and the test period more than the other groups. Interestingly enough, the rote-learning treatment group was consistently superior to the other groups. Kersh interprets this phenomenon as being a function of retroactive inhibition acting on the guided discovery and directed-learning treatment groups. Support was also claimed for the notion that self-discovery motivates the subject to practice more and thus remember and transfer more than he might if taught directly.

Kersh interpreted the results of the two experiments to be complimentary by stating that the direct reference group (1958 study) and the guided discovery (1962 study) groups were the same and that the rule given (1958 study) and rote learning (1962 study) groups corresponded. He claimed no counterparts for the "no help" treatment (1958 study) and for the directed learning treatment (1962 study).

In a series of studies completed at the University of Illinois, (Ray, 1957, Rowlett, 1960, Moss, 1960, Grote, 1960, Tomlinson, 1962) a number of researchers have studied the effects of directed discovery and direct-detailed methods of teaching complex technical material, drawn from the area of industrial education. In each case the direct detailed method of presentation consisted of a highly organized and detailed step by step procedure covering facts, principles and relationships. The directed discovery sequence, in contrast, consisted of leading questions and sequences of questions used to direct the subjects attention to problem situations, possible alternate meanings and applications, and the problem solutions which were to be discovered by the subjects.

Ray (1957) selected the use and understanding of the vernier micrometer caliper as his learning task. Subjects were ninth-grade boys whose mean intelligence score was near the mean of the test norm groups. Subjects were given a 40 minute learning period followed by a test on the material learned and were tested one week and six weeks after training. Differences on a test of retention favored the directed discovery treatment after six weeks although no significant differences were observed after one week. On a transfer task after both one week and six weeks, differences favored the directed discovery treatment. There were no significant interaction effects, but Ray did comment that, if anything, the directed discovery treatment was more appropriate at lower ability levels.

Rowlett (1960) applied the same direct-detailed, directed discovery methods to the learning of orthographic projection principles and skills. The design was similar to that of Ray (1957). Subjects were given a 41 minute learning program and a test based on materials learned. This was followed by tests of retention and transfer at twelve days and six weeks. Rowlett concluded that the groups did not differ initially, but that the directed discovery treatment was superior to the direct-detailed treatment on a test of retention and transfer at both twelve days and six weeks.

Moss (1960) repeated the above designs using a task of learning letter press imposition with junior and senior students in a vocational high school. The mean I.Q. of Moss's group was at the 34th percentile (considerably lower than those in the studies of Ray (1957) and Rowlett (1960)). Moss reported no differences between directed discovery and direct-detailed groups initially or after one week and six weeks. Neither did he find any evidence of a method by ability level interaction on objective criterion measures.

Grote (1960) studied high ability students in a modification of the above designs. He gave two groups of subjects instruction in the principles of mechanics applied to the lever by one of the two methods described above. Eight days later each group was divided into two sub-groups, each sub-group receiving instruction in the principles of mechanics applied to the wheel and axle by one of the two treatments listed above. Thus, there were four treatments: (a) directed discovery-directed discovery (BB), (b) directed discovery-direct detailed (BA); (c) direct detailed-direct detailed (AA); and (d) direct detailed-directed discovery (AB). Subjects were tested immediately after each instructional session and again at one week and six weeks after the second instructional period. After the first learning session the direct-detailed method (A) was superior to the directed discovery method (B). The only significant differences found for the method or sequence on retention and transfer problems, favored the BA sequence on both retention and transfer at one week. Grote also found that boys consistently performed better than girls on each criterion measure.

Gagné and Brown (1961) investigated the effect of presenting programed conceptual learning materials by three variations of method. The task involved the summing of number series. One treatment (RE-rule and example) received a verbal statement of the principle and formal practice in the application of each rule through a series of examples. The second method (GD-guided discovery) consisted of giving the subjects examples and guiding them through a series of directed questions to a final formulation of the rule. Subjects in the third method (D-discovery) were required to independently discover the rules from several different series of examples. The criterion measure was a transfer test given almost immediately after the learning period. The results showed the guided discovery method to be most effective in all cases, but only slightly superior to the discovery group. The discovery group took less time than either of the other groups, however. Both discovery methods were definitely superior to the rule and example treatment. Gagné and Brown interpreted their findings as being consistent with the findings of Katona (1940) and Haslerud and Meyers (1958). They concluded:

> However, our results emphasized the importance of "what is learned" rather than "how it is learned" as the crucial factor in learning effectiveness. Discovery as a method appears to gain its effectiveness from the fact that it requires the individual learner to reinstate (and in this sense to practice) the concepts he will later use in solving new problems (Gagné and Brown, 1961, p. 320).

Della-Piana and Eldredge (1964) questioned the interpretation of the Gagné-Brown study. They pointed out differences in the GD and RE treatments both in what was practiced (e.g., whether the rule for summing series was to be written) and how it was practiced (e.g., the prompting and fading conditions under which a response is called for).

Gagné and Smith (1962) investigated the effect of verbalization and solution set instructions on subjects involved in the three-disk, three-circle problem previously used by Ewert and Lambert (1932). The groups were differentially treated with respect to solution set (being told to look for a principle to be stated verbally) and verbalization during practice. Gagné and Smith found that verbalization made a significant difference, while solution set instructions made none. The subjects who were required to verbalize during practice were significantly superior, on the transfer test, to those who had not been required to verbalize. Gagne and Smith interpreted the results to show that requiring subjects to verbalize during practice facilitates discovery of the general principles as well as transfer.

Suchman (1962) matched experimental and control groups of fifth and sixth grade students in twelve elementary schools according to intelligence and ability. Science was taught by an inquiry training method to experimental groups, whereas control groups received classroom instruction. Inquiry Training was superior to expository teaching in its effect on conceptual growth, but there were no significant differences in its effect on the productivity of inquiry. The inquiry group had a significantly higher question-asking fluency score.

Blitz (1962) studied determinants of speed of problem solving. In two practice tasks subjects were given reinforcement which varied in degree of relevancy and number. It was found that the speed of solving the test problem was a direct function of the degree of relevancy of the prior reinforced relevant response. However, the variable "number of reinforcements" interacted with the variable "degree of relevancy" in such a way that the effect of degree of relevancy was more pronounced at the higher level of reinforcement.

Scandura (1962) studied problem solving by subjects taught by discovery (D) and expository (E) processes. He found that D subjects who were successful at solving routine problems were also successful at solving novel problems. This finding did not hold for E subjects. D subjects also looked for more ways to solve problems whereas E subjects tended to be more procedure oriented. E subjects made fewer mistakes on routine problems but had difficulty adapting to new problems and D subjects considered novel problems to be more of a challenge.

An experiment with 243 high school pupils (Ashton, 1962) compared a heuristic method (guided discovery) with a textbook method (expository method) on a problem solving task. The heuristic method proved superior to the textbook method, although the type or precise amount of guidance given to the heuristic treatment group was not clearly specified.

Wittrock (1963) used college students as subjects in an investigation of the relative effectiveness of four methods on a code-deciphering task. The four methods were: (a) rule given, answer given; (b) rule given, answer not given; (c) rule not given, answer given; and (d) rule not given, answer not given. On a test of initial learning, those groups which received the rules did significantly better than those groups to which the rules were not given. The same was true with reference to a retention and transfer test. The discovery group (rule not given, answer not given) was the only group which exhibited retention scores higher than the initial learning scores, however.

A total of 454 eighth grade students were randomly divided into four experimental groups and a control group in a study by Hountras and Belcastro (1963). They attempted to determine the relative effectiveness of four techniques of programming the addition and subtraction axioms of algebra; namely, the non-verbal inductive technique, the verbal inductive technique, the non-verbal deductive technique, and the verbal deductive technique. All groups were equated for intelligence, age, achievement in reading and arithmetic, and socio-economic level. Each program presented a sequence of materials which was designed to lead the pupil gradually from simple to complex material and total mastery. Each program required some response on the part of the pupil and also provided immediate knowledge of results. In every set of frames a progression of prompts or cues was also provided so that maximal cueing occurred initially and minimal cueing occurred in the final stages of the sequence. The control group did not receive any instruction. All groups were pre- and post-tested. Analysis of variance revealed significant differences among the four techniques of programming algebraic axioms. The verbal deductive technique was superior to all other techniques. Further, experimental groups performed better than the control group. On the whole, girls were better than boys in learning from programed materials.

Yabroff (1963) conducted an experiment with 272 college students to determine the effects of inductive and deductive teaching methods on the speed and accuracy of relevant cognitive activities and on interest factors. Two arrangements of each method were used. In the blocked arrangement, all the rules and definitions for each concept were grouped together, and all the sample problems about that concept were grouped together. In the mixed arrangement, rules and sample problems for a given concept were interspersed, providing frequent shifts from examples to rules and back again. Millers Analogies Scores, Form A, were used to separate the sample into high and low intelligence levels. Subjects were assigned at random to each arrangement of the inductive and deductive method. Linear programed instructional material in booklet form was used to teach each treatment group. The subject matter programed consisted of selected concepts from the field of elementary statistics and test interpretation. There were no significant differences in scores on the transfer test or on the time required to complete the programed booklet. However, the inductive method produced greater speed in transfer of rules and the deductive method generated a more favorable attitude. No significant differences existed between mixed and blocked arrangements of each method on any criterion. Intelligence has definite effects on performance.

Blank (1963) trained children by means of auto-instructional procedures to ask questions about the relevant dimensions of problems before attempting to solve them. The subjects, 54 sixth grade pupils enrolled in summer session science classes, were divided into three groups matched for I.Q. and known tendency to ask questions. Students in the first experimental group studied the problems by means of the auto-instructional unit; those in the second experimental group read the same problems, presented in programmed format but with the inquiry training material deleted; those in the control group saw neither the program nor any of the problems. All pupils were tested and retested orally on a set of criterion problems by trained interviewers. In addition, they respond in writing to another criterion test and to a teacher-made achievement test. Following the post-tests, each pupil was rated on the extent of his participation in class discussion. Analysis of the data showed that the children given the inquiry training program asked significantly more relevant questions on both oral and written criterion tests and obtained higher ratings on class participation.

Wittrock (1964) experimented further with the types of cues given to subjects working on an abstract mathematical learning task. His subjects were divided into four groups: (a) a group which received only general cues, (b) a group which received cues pertaining to various classes of concepts, (c) a group which received specific cues, and (d) a group which received no cues. The results indicated that class cues were more efficient on a retention and transfer criterion measure, while general cues were least efficient.

In a study by Blackman and Holden (1964) two groups of educable, non-reading retarded adolescents were compared on performance in a support (prompting) and a non-support (confirmation) program designed to teach 4 words. Each group contained 12 subjects equated on the basis of IQ and previous performance. Programs were presented on a teaching machine, with a button press and multiple-choice response. Data analyses of number of responses required to complete 16 test frames and number of words successfully read from cards immediately after and 24 hours after training showed no significant differences between groups. IQ and number of responses required to complete the program showed an inverse relationship, but this relationship was not significantly different for the two groups. Variations within the support program, involving full support, reduced support, and test frame, had no effect on the mean number of responses.

Kersh (1965) tested the dual hypothesis that discovery learning intensifies the "searching behavior" of subjects and that subjects taught by discovery techniques should maintain their initial performance through post-experimental practice. Subjects used were fifth-grade pupils and the learning task employed was that of programed instructional sequences designed to teach the distributive law of multiplication over addition. Three treatments were used: (a) programed discovery, (b) programed guidance, and (c) free discovery. Kersh found no differences among the groups on retention tests given three days, two weeks, and six weeks after the end of the learning period. On an oral problem-solving test given 24 hours after the end of the instructional period, Kersh found that the programed discovery group evidenced less of a tendency to search for patterns than evidenced in either of the other treatment groups. In addition, Kersh's earlier findings (1958, 1962) concerning the motivational effect of discovery learning, as evidenced in post-experimental practice, were not supported by this study. In this experiment there were no differences in the post-experimental practice reported by subjects in the various groups. Kersh interpreted this finding as being partially accounted for by the length and lack of continuity of the learning experiences. (The subject's task application was for relatively short periods, once a day, over a relatively long period of time.) Kersh concluded that such interspersed tasks may actually be more representative of classroom conditions, and hence more valid in measuring actual "practical" advantages or disadvantages of discovery learning than experiments of shorter duration.

Wolfe (1963) reported a study in which 300 pupils who were accustomed to discovery methods (having had one year of mathematics in the University of Illinois Committee on School Mathematics program) were taught programed mathematical material by an expository method and a discovery method. This study was designed to investigate possible effects that presenting mathematical subject matter by an expository method may have on pupils who were primarily accustomed to learning mathematical subject matter by discovery techniques. The pupils used in this study were divided into low, average, and high ability sub-groups at both the ninth- and tenth-grade level. Wolfe found that the two treatments used in his study produced no significant differences at any ability or grade level, nor were there any significant interaction effects. He concluded that a change to expository instruction for pupils accustomed to being taught by a discovery method did not seem to interfere with their ability to learn mathematics. He speculated, however, that, since pupils in the expository group had previously been trained in discovery techniques, their understanding might possibly have been gained by discovering from the examples rather than from the verbal statement of the generalization given in the expository treatment.

Twelker (1965) used secondary school students to test the hypothesis that "feedback which included rules and answers enhances performance on tests of learning, retention, and transfer to new problems of the same class as that taught to a greater extent when rules as prompts are not given to learners than when rules as prompts are given to learners".

Twelker used a twenty frame program teaching some elements of cryptography. All subjects received the same initial instruction. After eight introductory frames, subjects were divided into eight treatments based upon combinations of three variables: (a) rule prompts given or not given; (b) answer prompts - given or not given; and (c) feedback - given or not given.

Subjects were tested immediately and one week later on tests of retention and transfer. Results indicated that on an immediate test of learning, giving rules as compared with withholding rules enhanced performance for students of both high and low achievement. Interaction effects indicated that giving answers to specific problems, or feedback, was more profitable to students of low achievement level than students of high achievement level. Interaction effects also indicated that feedback enhanced learning to a greater extent when answer prompts were not given than when prompts were given. On an immediate test of transfer to new examples of the learned rules and new transpositional rules, it was shown that giving rules produced reliably higher scores than withholding rules. In contrast to the hypothesis concerning feedback, feedback enhanced performance only on the delayed test of learning.

Summary of the Literature. The noncomparability of the various studies discourages any serious attempt to summarize the results with any degree of confidence. Perhaps the best that can be dome is a simple arithmetic count of the results from studies which compare "meaningful" vs. "rote" methods. Such a count would favor "meaningful" over "rote" methods at a five to one ratio.

Unfortunately, the other dimensions of discovery are not so easily counted and compared. The precise method by which the learners were taught is not clear in many of the "discovery" studies. The difficulty in attempting to summarize such research was recognized by Henderson (1963), who, when attempting to review discovery research as it pertained to math, made the statement:

> Before turning to other kinds of research on teaching secondary school mathematics, let us consider what can be concluded about the consequences od tell-and-do methods versus heuristic methods. One conclusion is that the evidence is not conclusive . . .
>
> One is tempted to admonish the reader to draw his own conclusions about the findings and conclusions of the various experimenters. But more than this can be offered even if it is negative in tone. If the consequences of being wrong are serious, one should be wary of adopting one method rather than another on the basis of the evidence available. If the consequences of being wrong are not serious, one is freer to choose (Henderson, 1963, pp. 1020-1021).

A review of the literature on discovery learning indicates that there is no unequivocal evidence for the superiority of any method defined along the discovery-expository continuum. Rather, the findings, at the present time, have been inconclusive. It is generally agreed by those who have reviewed the research literature pertinent to the discovery-expository dilemma that more adequate control of relevant variables is ncessary. Only with such control can any determination be made of the relative efficacy of one teaching method as compared with another. As Ausubel notes:

> In many cases, of course, findings are equivocal simply because of failure to control such other relevant variables as the rote-meaningful, the inductive-deductive, the verbal-nonverbal, and the intramaterial organization dimensions of learning, while varying the reception-discovery factor. In other instances, it is quite possible that negative findings are less indicative of inadequacies in the underlying theory than of inadequacies in research design, which unfairly load the dice against the possibility of confirming hypotheses (Ausubel, 1961b, pp. 47-48).

A recent paper by Cronbach (1965) specifies twelve different outcomes claimed by spokesmen for discovery methods and calls for research on interactions among subject matter, type of instruction, timing of instruction, type of pupil and outcome.

The research also reveals that almost all of the experiments which have been conducted have been so limited in scope that one might readily question any attempt to generalize the results to the classroom situation. Most "discovery" studies have dealt with extremely small time samples, small number of subjects, or both. The majority of experimenters have utilized learning tasks such as disk arrangement, card sorting, matchstick manipulation, use of nonsense syllables or scrambled alphabets, etc. At best, these represent samples of limited, easily controlled tasks which can be approached by varying methods of instruction. At worst, they represent tasks so restrictive that one might argue they are scarcely related, even remotely, to anything which occurs in the typical classroom. While fully acknowledging the appropriate contribution of short-term laboratory studies, it is possible to conduct a carefully-controlled classroom experiment where both time sample and learning task are representative of typical school behavior and curriculum. Certainly the results of such a study could be generalized to the classroom with more confidence than those of the typical short-term laboratory experiment.

Hilgard seems to support this point of view when he argues that researchers should become immersed in "real life" data as well as nonsense syllables, mazes, target practice, etc., (Hilgard, 1956, pp. 480-481).

Berelson and Steiner note, with reference to laboratory investigation, ". . . relatively little of what psychologists study under this heading deals with the kind of learning that occurs in the classroom . . ." (Berelson and Steiner, 1964, p. 133). They go on to suggest that:

> . . . laboratory studies of human learning have concentrated on the relatively simple mechanical processes by which people acquire skills or memorize lists of items. The learning of ideas or the attainment of understanding . . . has received less systematic attention . . . (Berelson and Steiner, 1964, p. 134).

In the studies reported here we have used a relatively simple short-term self-instructional sequence (Eldredge's study) and a more extensive time sample (Worthen's study) utilizing six weeks of mathematics instruction including teacher behavior with well structured instructional sequences.

References

Ackerman, W. I., *Presentation of Alternatives and its Relation to Set in Problem Solving*, Ed. D. Dissertation, Cambridge, Massachusetts: Harvard University, 1956, 146 Pp.

Anderson, G. L., "Quantitative Thinking as Developed Under Connectionist and Field Theories of Learning," *Learning Theory in School Situations*. University of Minnesota Studies in Education No. 2, Minneapolis: The University of Minnesota Press, 1949, 40-73.

Ashton, Sister Madeleine, R., "Heuristic Methods in Problem Solving in Ninth Grade Algebra," *Dissertation Abstracts*, 1962, 62, 2320.

Ausubel, D. P., "In Defense of Verbal Learning," *Educational Theory*, January, 1961, XI (1), 15-25. (a)

Ausubel, D. P., "Learning by Discovery: Rationale and Mystique," *The Bulletin of the National Association of Secondary School Principals*, December, 1961, 45 (269), 18-58. (b)

Ausubel, D. P., *The Psychology of Meaningful Verbal Learning*. New York and London: Grune and Stratton, 1963. (b)

Ausubel, D. P., "Some Psychological and Educational Limitations of Learning by Discovery," *The Arithmetic Teacher*, May, 1964, 11, (5), 290-302,

Ballard, MarJean L., "An Experimental Study of Discovery Learning," Unpublished Master's Thesis, University of Utah, 1964.

Beberman, M., *An Emerging Program of Secondary School Mathematics*. Inglis Lecture Series, Cambridge: Harvard University Press, 1958.

Berelson, B., and Steiner, G. A., *Human Behavior, An Inventory of Scientific Findings*. New York: Harcourt, Brace, and World, Inc., 1964.

Blackman, L. S. and Holden, E. A., Jr., "Support vs. Non-Support in an Auto-instructional Word Program for Educable Retardates," Chapter 12 in *The Development and Evaluation of a Curriculum for Educable Mental Retardates Utilizing Self-Instructor Devices or Teaching Machines*, NDEA Title VII Project No. 368, Washington, D. C.: U. S. Office of Education, February, 1964, 190-200.

Blank, S. S., "Inquiry Training Through Programmed Instruction," *Dissertation Abstracts*, September, 1963, 24 (3), 1071. (Abstract: Ph.D. Dissertation, University of California, Berkeley, 1963.)

Blitz, B., "Relevant Responses in Problem Solving: The Role of Degree of Relevancy, Number of Reinforcements and Super-Ordinate Responses," *Dissertation Abstracts*, July, 1963, 24 (1), 396-397. (Abstract: Ph.D. Dissertation, New York University, 1962.)

Bruner, J. S., "The Act of Discovery," New Frontiers in Reading, Figurel, J. A. (Ed.), Scholastic Magazines, New York, 1960.

Corman, B. R., "The Effect of Varying Amounts and Kinds of Information as Guidance in Problem Solving," Psychological Monographs, 1957, 71 (2).

Craig, R. C., The Transfer Value of Guided Learning. New York: Bureau of Publications, Teachers College, Columbia University, 1953.

Craig, R. C., "Directed vs. Independent Discovery of Established Relationships," Journal of Educational Psychology, 1956, 47, 223-24.

Cronbach, L. J., "The Logic of Experiments on Discovery," Mimeo. Paper prepared for a conference in New York City, January, 1965.

Davis, R. B., Discovery in Algebra. A Workbook in Introductory Mathematics Developed from the Madison Project Experimentation, Massachusetts: Addison Wesley, Inc., 1960. (a)

Davis, R. B., Teacher's Manual for Discovery in Algebra: A Workbook in Introductory Mathematics Developed from the Madison Project Experimentation. Massachusetts: Addison Wesley, Inc., 1960. (b)

Davis, R. B., A Modern Mathematics Program as it Pertains to the Interrelationship of Mathematical Content, Teaching Methods and Classroom Atmosphere. (The Madison Project) Syracuse University, Cooperative Research Project No. D-022, 1962.

Davis, R. B., Elementary Mathematics Series: Axioms for Arithmetic and Algebra. Madison Project, Syracuse University, Webster College, 1963.

Davis, R. B., Gleason, A. M., Lomon, E. L., Moise, E. E., and Springer, G., "Goals for School Mathematics: Report of the Cambridge Conference on School Mathematics," (Mimeo first draft of Final Report), August, 1963.

Della-Piana, G. M., "Searching Orientation and Concept Learning," Journal of Educational Psychology, 1957, 48, 245-253.

Della-Piana, G. M., and Eldredge, G., "Discovery Learning in Programed Instruction," Paper presented at the Second Annual Convention of the National Society for Programmed Instruction, San Antonio, Texas, April 3, 1964.

Ewert, P. H., and Lambert, J. F., "Part II: The Effect of Verbal Instructions Upon the Formation of a Concept," Journal of General Psychology, 1932, 6, 400-413.

Fehr, H. F., Teaching High School Mathematics: What Research Says to the Teacher. National Education Association, Washington, D. C., 1961.

Forgus, R. H., and Schwartz, R. J., "Efficient Retention and Transfer as Affected by Learning Method," Journal of Psychology, 1957, 43, 135-139.

Fullerton, C. K., "A Comparison of the Effectiveness of two Prescribed Methods of Teaching Multiplication of Whole Numbers," Dissertation Abstracts, 1955, 15, 2126-2127.

Gagné, R. M., and Brown, L. T., "Some Factors in the Programing of Conceptual Learning," Journal of Experimental Psychology, 1961, 62, 313-321.

Gagné, R. M., and Smith, E. C., Jr., "A Study of the Effects of Verbalization on Problem Solving," Journal of Experimental Psychology, 1962, 63 (1), 12-18.

Grote, C. N., "A Comparison of the Relative Effectiveness of Direct-detailed and Directed Discovery Methods of Teaching Selected Principles of Mechanics in the Area of Physics," Unpublished Ed.D. Dissertation, Urbana: University of Illinois, 1960.

Haslerud, G. M., and Meyers, Shirley, "The Transfer Value of Given and Individually Derived Principles," Journal of Educational Psychology, December, 1958, 293-298.

Henderson, K. B., "Research on Teaching Secondary School Mathematics," Handbook of Research on Teaching, Gage, N. L. (Ed.) Chicago: Rand McNally, 1963, 1020-1021.

Hendrix, Gertrude, "A New Clue to Transfer of Training," Elementary School Journal, December, 1947, 48, 197-208.

Hendrix, Gertrude, "Learning by Discovery," The Mathematics Teacher, May, 1961, 54 (5), 290-299.

Hilgard, E. R., Theories of Learning. New York: Appleton-Century-Crofts, Inc., 1956.

Hountras, P. T. and Belcastro, F. P., "A Comparison of Four Techniques of Programming Algebra," Science Education, October, 1963, 47 (4), 388-393.

Judson, A. J., Cofer, C. N., and Gelfand, S., "Reasoning as an Associative Process: II. 'Direction' in Problem Solving as a Function of Prior Reinforcement of Relevant Responses," Psychological Reports, December, 1956, 2 (4), 501-507.

Katona, G., Organizing and Memorizing. New York: Columbia University Press, 1940.

Kersh, B. Y., "The Adequacy of 'Meaning' as an Explanation for the Superiority of Learning by Independent Discovery," Journal of Educational Psychology, 1958, 49 (5), 282-292.

Kersh, B. Y., "The Motivating Effect of Learning by Directed Discovery," Journal of Educational Psychology, 1962, 53 (2), 65-71.

Kersh, B. Y., "Learning by Discovery: What is Learned?" Speech read at twenty-third Summer Meeting of the National Council of Teachers of Mathematics, University of Oregon, Eugene, Oregon, August 22, 1963.

Kersh, B. Y., "Directed Discovery vs. Programed Instruction: A Test of a Theoretical Position Involving Education Technology," Paper read at American Educational Research Association meetings, Chicago, Illinois, February 12, 1965.

Kersh, B. Y. and Wittrock, M. C., "Learning by Discovery: An Interpretation of Recent Research," Journal of Teacher Education, In press, 1962.

Kittell, J. E., "An Experimental Study of the Effect of External Direction During Learning on Transfer and Retention of Principles," Journal of Educational Psychology, 1957, 48, 391-405.

McConnell, T, R., "Discovery vs. Authoritative Identification in the Learning of Children," Studies in the Psychology of Learning II, Iowa City, Iowa: University of Iowa Studies in Education, 1934, 5, 13-62.

McCoy, M. Eleanor, "A Secondary School Mathematics Program," The Bulletin of the National Association of Secondary School Principals, May, 1959, 12-18.

Morgan, J. J. B., "Value of Wrong Responses in Inductive Reasoning," Psychological Abstracts, 1945, 19, 241. (Abstract: Journal of Experimental Psychology, 1945, 35, 141-146.)

Moss, J., Jr., "An Experimental Study of the Relative Effectiveness of the Direct-detailed and Directed Discovery Methods of Teaching Letterpress Imposition," Unpublished Ed.D. Dissertation, Urbana: University of Illinois, 1960.

Norman, Martha, "Three Methods of Teaching Basic Division Facts," Dissertation Abstracts, 1955, 15, 2134.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H., The Measurement of Meaning. Urbana; University of Illinois Press, 1958.

Petronia, Sister Mary, S. S. N. D., "The Pros and Cons of Modern Elementary Mathematics," National Catholic Educational Association Bulletin, August, 1962, 381-382.

Pincus, M., "An Investigation into the Effectiveness of Two Methods of Instruction in Addition and Subtraction Facts," Dissertation Abstracts, 1956, 16, 1415.

Ray, W. S., "Pupil Discovery vs. Direct Instruction," Journal of Experimental Education, March, 1961, 29 (3), 271-280.

Rowlett, J. D., "An Experimental Comparison of Direct-detailed and Directed Discovery Methods of Teaching Orthographic Projection Principles and Skills," Unpublished Ed.D. Dissertation, Urbana: University of Illinois, 1960.

Scandura, J. M., "The Teaching-Learning Process: An Exploratory Investigation of Expository and Discovery Modes of Problem Solving Instruction," Unpublished doctoral dissertation, Syracuse University, 1962.

Schwartz, B. D., "Certain Relationships Between Verbalization and Concept Formation: The Destruction of Ideas by Words," Unpublished doctoral dissertation, Princeton University, 1948.

Shaw, J. A., "A Students' Inventory of Teaching-Learning Processes: A Manual for Forms C and D," (Mimeo) Colorado State College, 1963.

Sobel, M. A., "A Comparison of Two Methods of Teaching Certain Topics in Ninth Grade Algebra," Dissertation Abstracts, 1954, 14, 1647.

Stacey, C. L., "The Law of Effect in the Retained Situation with Meaningful Materials," Learning Theory in School Situations, University of Minnesota Studies in Education, No. 2, Minneapolis: The University of Minnesota Press, 1949, 74-103.

Suchman, J. R., The Elementary School Training Program in Scientific Inquiry. University of Illinois, 1962. Title VII Project No. 216 (U. S. Office of Education, Washington, D. C., June, 1962), 128 pp.

Swenson, Esther, J., "Organization and Generalization as Factors in Learning, Transfer, and Retroactive Inhibition," Learning Theory in School Situations. University of Minnesota Studies in Education No. 2, Minneapolis: The University of Minnesota Press, 1949, 9-39.

Thiele, C. L., The Contribution of Generalization to the Learning of the Addition Facts. New York: Columbia University Press, 1938.

Tomlinson, R. M., "A Comparison of Four Methods of Presentation for Teaching Complex Technical Material," Unpublished Ed.D. Dissertation, Urbana; University of Illinois, 1962.

Twelker, P. A., "Rules, Answers, and Feedback in Learning Retention and Transfer of Concepts," Paper read at the American Educational Research Association, Chicago: February, 1965.

Umemto, T. and Haslerud, G. "Responses and Attitudes of American and Japanese Children to Number Situations: A Comparative Study," Originally printed in Japanese, 1955, translated by Tatsuoko, M. M., University of Illinois,

Willoughby, S. S., "Discovery," The Mathematics Teacher, January, 1963, 22-25.

Wirtz, R. W., Botel, M., and Nunley, B. G., Discovery in Elementary School Mathematics: Classroom Training Units for Teachers of Modern Mathematics. Chicago: Encyclopaedia Britannica Press, Inc., 1963.

Wittrock, M. C., "Verbal Stimuli in Concept Formation: Learning by Discovery," Journal of Educational Psychology, 1963, 54, 183-190.

Wittrock, M. C., "Transfer Through Verbal Cueing in Concept Identification," Cooperative Research Project No. 1684, Los Angeles: University of California, 1964.

Wolfe, M. S., "Effects of Expository Instruction in Mathematics on Students Accustomed to Discovery Methods," Unpublished doctoral dissertation, University of Illinois, 1963.

Yabroff, W. W., "The Comparative Effects of Inductive and Deductive Sequences in Programed Instruction," Dissertation Abstracts, February, 1964, 24:8, 3214-3215. (Abstract: Ph.D. Dissertation, Stanford University, 1963.)

CHAPTER III

TASK ANALYSIS*

G. Della-Piana and Garth Eldredge

There is a growing interest among educators, psychologists and subject matter experts in the problem of finding terminology and procedures for specifying structure and sequence parameters of learning tasks. Until we can describe the structure and sequence of text materials in ways that are interpretable and that capture the significant dimensions relevant to learning outcomes, we are not likely to develop much generalizable knowledge because of the difficulty of replicating studies in the task parameters are ambiguous and the futility of replicating studies where the task parameters do not account for much of the variability in learning outcome. The search for task analysis procedures in so many quarters is ample evidence of the awareness of the significance of this problem: Ausubel (1963, Chapter 8), Ausubel (1964), Evans, Glaser and Homme (1960), Gagné (1962), Glaser (1962, pp. 8, 57, 178-84, 436-37, 483), Stolurow (1961, pp. 64-73), Schwab (1964a), Schwab (1964b), Woodruff (in press). In this chapter we shall not attempt a review of these pioneering efforts. Instead we present first an illustration of the problems of controlling relevant nonexperimental variables without an adequate task analysis and then a description of the structure and sequence of the tasks used in our experiments. A discussion of how such controls partially destroy the treatment or phenomena the experimenter may be most interested in is discussed in Chapter I.

---

*This chapter is a modification and extension of a paper on "Discovery Learning and Programed Instruction" presented by the authors at the Second Annual Convention of the National Society for Programmed Instruction, San Antonio, Texas, April 3, 1964. A description of the controls in the treatments of Worthen's study is presented separately in Chapter V.

An Illustration of the Problem

The difficulties inherent in interpreting control group studies of so-called "discovery learning" may be illustrated by an intensive analysis of the Guided Discovery (GD) and Rule and Example (RE) learning programs of Gagné and Brown (1961).

A summary of the notation used in describing the programs and a sample of frames from the original Gagné-Brown GD and RE programs are presented in Exhibits A, B, and C respectively at the end of this chapter. The frames reproduced are parallel frames from the GD and RE programs in which one rule is taught for summing the number series 1, 2, 4, 8, 16 . . . The reader should study Exhibits A, B, and C before proceeding with the following text.

Similarities in the GD and RE Programs.--

1. The total formula is called for *once* in each program in the symbolic notation form, $\overset{n}{\Sigma} = T_{n+1} - 1$. (See RE program frame 91, GD program frame 95.)

2. The numerical value of the sum, $\underline{\Sigma}$, of the number series for a given *n*, or term number, is called for three times in each program. (See RE frames 95, 96, 97, and GD frames 96, 97, 98.)

3. The appearance of the same number series in each frame with the term number, *n*, above each term value, *T*, of the number series and the cumulative sum, $\underline{\Sigma}$, below each term value of the series is a feature occurring in each frame in almost completely identical form in both programs.

Differences Between the GD and RE Programs.--

1. The response of giving the term number, *n*, for a given sum,

$\underline{\Sigma}$, is never called for in the GD program and called for twice (frames 99, 100) in the RE program.

2. Giving part of the rule for summing the series is called for twice in the GD program (frame 91 "subtracting one" and frame 93 "T_{n+1} or the next following T") and never in the RE program.

3. The numerical value of T_{n+1} for different values of n is called for four times in the RE program (frames 92, 93, 94, 99) as compared with twice in the GD program (frames 92, 94).

4. Responses in all ten frames of the RE program sequence are called for under the prompt of a full statement of the formula in symbolic form ($\Sigma = T_{n+1} - 1$) but in the GD program only one frame (#96) contains this prompt. On the other hand a verbal statement of the formula (with or without symbols) is present only once in the RE program as a prompt (frame 97) and three times (frames 91, 92, 93) in the GD program. In other words the responses called for in the RE program are more highly prompted throughout and are under conditions of prompts with less associative value than in the GD program.

5. The response of producing the numerical sum, Σ, of the series for a given n is called for in the GD program under conditions of fading prompts but in the RE program under the conditions of increasing prompts.

6. The response of producing the numerical value of T_{n+1} is also called for under more highly prompted conditions in the RE program than in the GD program.

Interpretability of Treatment Differences.--

Gagné and Brown found GD to be superior to RE in facilitating summing of a new (untaught) number series. To what can we attribute these observed differences in treatment effectiveness? The above task analysis has clearly demonstrated that the GD and RE programs differ not only in what was practiced but also in how (prompting and fading conditions) it was practiced. It is, of course, true that some differences in what is practiced and how it is practiced are essential distinguishing characteristics of "discovery" as compared with"direct exposition" methods. However, some of the advantages given to the GD program in association value or meaningfulness as well as in fading of prompts could be part of an RE program without making it a GD program. An adequate test of the superiority of some form of guided discovery over direct exposition requires controlling relevant nonexperimental variables. In the final two sections of this chapter we describe how we attempted to specify and control relevant nonexperimental variables while varying the crucial experimental variables for the studies reported in Chapters IV and V.

Development of the Programed Unit on Summing Number Series

Having demonstrated the lack of control of relevant nonexperimental variables in the Gagné-Brown GD and RE programs the next task was to revise the programs to overcome these limitations. This task could not be achieved without some sacrifice in the direction of destroying some of the phenomena (treatment characteristics) we were most interested in investigating. A discussion of the problem of "destroying the treatment" is presented in Chapter I. Here we shall simply describe our instructional sequences. To allow a comparison of our revision with the original program we have included in Exhibits D and E at the conclusion of this chapter the revised version of the original program samples presented in Exhibits B and C. A listing of similarities and differences between the revised GD and RE programs in structural and sequence characteristics will clarify the nature of our experimental variable and the adequacy of our control of nonexperimental relevant variables. This description should be contrasted with the above description of the original programs. The listing of similarities between the revised GD and RE programs will clarify how nonexperimental relevant variables have been controlled. The listing of differences between a sample of the revised GD and RE programs will clarify exactly what our experimental variable (loosely called (discovery") really is in terms of structure and sequence characteristics of the text materials. A complete copy of the revised GD and RE programs is presented in Part II of the final report, published in a separate volume.

Similarities in the Revised GD and RE Programs.--

1. The total formula is called for three times in symbolic notation and four times in words in both programs.

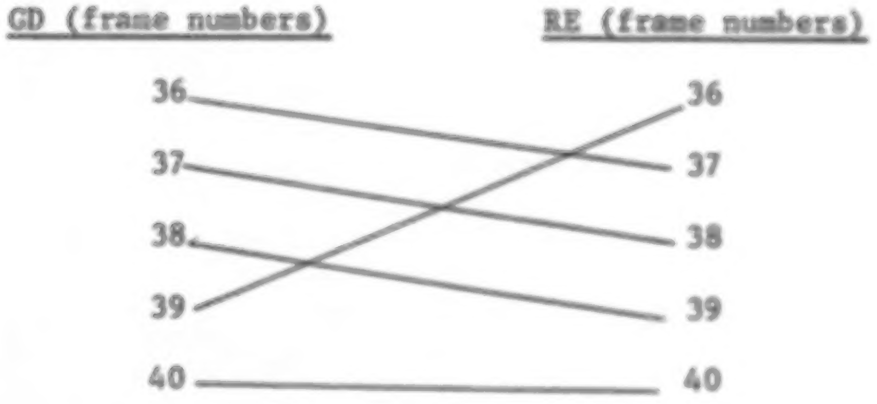
2. The numerical value of the sum, Σ, of the number series for a given n, or term number, is called for 11 times in each program. The value of n for each sum called for was also the same in each program.

3. The appearance of the same number series in each frame in the form ($S_{nT\Sigma}$) illustrated in Exhibit A occurred essentially the same in each frame of each program.

4. The number of times that a criterion response is called for under unprompted conditions is the same for both treatments unlike the original programs in which the RE treatment was consistently more highly prompted throughout the sequence.

Differences Between the Revised GD and RE programs.--

The major difference is that of sequence. The following parallelism of frames may be noted. Thus, each GD frame has a counterpart that is



equivalent (sometimes identical) in the RE program.

The major distinction between the two programs is in sequencing and in the prompting conditions under which the rule for summing the number series is first called for. In the GD program the rule is first called for under the prompt of directing the student's attention to the relationship between the sum and the term value of the $n + 1^{th}$ term. In the RE or expository program the rule is first called for under the prompt of a complete statement of the rule in mathematical notation and in words. But in both programs, the student at some time has an opportunity to state the rule and complete examples under highly prompted conditions.

Summary.--

In this chapter we have demonstrated how two treatments presumably equivalent on relevant nonexperimental variables were in fact grossly uncontrolled in this respect. We illustrated with a sequence of frames from our revision just how these variables were controlled. We ended up with a form of Expository instruction as follows:

1. Begin with a copying frame (copy the rule in symbols and words) and figure a sum with an example prompt).

2. Fade away the rules but not the examples in the next two frames while calling for the rule (in any form) and the sum of a given n terms in the series in one frame and the rule (in both forms) and the sum of a

given n terms in the series in the second frame.

3. Then call for the sum of a different n terms in the series under the prompt of the rule in mathematical notation.

4. Finally, call for the rule in words and in mathematical notation and call for the sum of n terms in the series (given n, or T_n or T_{n+1}) with no prompts.

The form of Guided Discovery we ended up with was largely equated with the Expository Treatment on relevant nonexperimental variables (although a few minor variations crept in) and can be summarized as follows:

1. Begin with a frame in which attention is called to possible relationships between the sum of n terms in the series and the term value of the n + 1th term and the learner is called upon to try to formulate a rule for this relationship and read the sum of n terms in the series from a table.

2. In the next frame the learner is asked to give the sum of n terms in the series under the prompt of a completed example. He is also asked to attempt a more complete formulation of the rule using mathematical notation (T_{n+1}) for "the term value of the n + 1th term in the series.

3. Then call for the sum of a different n terms in the series under the prompt of the rule given in mathematical notation.

4. Next call for copying of the rule in symbols and words and figuring a sum without an example prompt.

5. Finally, call for the rule in words and in mathematical notation and call for the sum of n terms in the series (given n, or T_n, or T_{n+1}) with no prompts.

Through largely controlling relevant nonexperimental variables we have enhanced the interpretability of our studies. Thus, treatment differences may fairly confidently be attributed to the sequence characteristics of the programs we have experimentally varied. However, in order to achieve this control it was necessary to destroy the "discovery" character of the GD treatment to some extent. A discussion of that problem, however, is contained in Chapter I of this report.

The structure of the programing for the Guided Discovery and Expository treatments of Worthen's study is presented in some detail in Chapter V of this report. Also, a complete reproduction of his materials is presented in Part II of this final report, published as a separate volume. The program discussed in this chapter was used in Eldredge's study reported in Chapter IV of this report. It is presented also in Part II of the final report; however, in the study it was printed on 5 x 7 cards, bound with rings, with answers on the backs of the cards.

Thus, see Chapter V for a rather detailed description of the materials and teaching procedures characteristic of "discovery" and "expository" classroom instruction as these treatments are implemented in Worthen's study. Rather than go into a detailed "task analysis" of those materials in the present chapter we shall simply list here the five dimensions on which expository and discovery teaching were differentiated in his study. Here we describe the "discovery" dimension leaving the definitions of expository to the imagination of the reader or to detailed treatment in Chapter V.

1. Teacher omniscience: In the discovery method the teacher avoids projecting the attitude that he knows all in order to deter students from looking to him for all answers or questioning the correctness of the teacher's response. At the same time, of course, the teacher must not appear uninformed.

2. Introduction of a generalization: In the discovery method the teacher delays verbalization of the concept until a minimum of 85 percent of the pupils have made the discovery.

3. Method of answering questions: When a question is asked in the discovery treatment, the teacher answers it by referring the pupil back to the material perhaps regrouping examples to facilitate discovery, but not verbalizing the principle.

4. Control of pupil interaction: In the discovery treatment an attempt is made to present students from prematurely revealing to other students discovery of a principle.

5. Method of eliminating false concepts: In both treatments a teacher includes "trap questions" which resemble problems which can be readily solved by a particular principle. In the discovery treatment, however, the teacher does not tell the student to watch for the trap questions nor show the student why the rule does not work after he errs on a trap question.

EXHIBIT A

NOTATION USED IN ANALYSIS OF PROGRAM FRAME COMPONENTS AND SEQUENCE CHARACTERISTICS

$S_{n,\,T,\,\Sigma}$ = a number series (S) including three rows: Term numbers (n), term values (T), and cumulative sums (Σ). In the program sample referred to in this chapter the series referred to is:

(n) Term number	1	2	3	4	5	6	7	8
(T) Number series	1	2	4	8	16	32	64	—
(Σ) Sums	1	3	7	15	31	63	—	—

T = a term value in the number series (T = 1, 2, 4, 8, 16, 32, 64 . . .)

n = the ordinal position of a given term value in the number series (n = 1, 2, 3, 4, . . .)

$\overset{n}{\Sigma}$ = the cumulative sum of n consecutive terms in the series (thus, when $n = 3$, $\overset{n}{\Sigma} = 7$; when $n = 5$, $\overset{n}{\Sigma} = 31$)

T_n = term value of n^{th} term (thus, when $n = 5$, $T_n = 16$)

T_{n+1} = term value of $n + 1^{th}$ term (thus, when $n = 5$, $T_{n+1} = 32$)

$\overset{n}{\Sigma}$ = $T_{n+1} - 1$ is one formula for summing the number series presented here in symbolic notation. (Thus, using the formula, if $n = 5$, $T_{n+1} = 32$, one less than 32 is 31 and thus the sum of the first 5 term values is 31.

2

EXHIBIT B

Frames 91 to 93 in the GD (Guided Discovery) Program in which the rule $\overset{n}{\Sigma} = T_{n+1} - 1$ is taught for summing the number series 1, 2, 4, 8, 16, . . .

91.

Term No.:	1	2	3	4	5	6	7	8
Series :	1	2	4	8	16	32	64	___
	↗	↗	↗	↗	↗	↗		
Σ :	1	3	7	15	31	63	___	___

Here is the same series with the three arrows written out. What do you have to do to obtain the sum, or $\overset{n}{\Sigma}$? Can you get $\overset{n}{\Sigma}$ by adding, subtracting, dividing, or multiplying the numbers in the row pointed to? For example, when $\overset{n}{\Sigma}$ is 3 (bottom row), the term value pointed to is 4. If you know 4 you can get 3 by _____ _____?

(Subtracting 1)

92.

Term No.:	1	2	3	4	5	6	7	8
Series :	1	2	4	8	16	32	64	___
	↗	↗	↗	↗	↗	↗		
Σ :	1	3	7	15	31	63	___	___

When $\overset{n}{\Sigma}$ is 3, this value can be obtained by subtracting 1 from next following T. When the $\overset{n}{\Sigma}$ is 7, this value can be obtained by subtracting 1 from the next following T, which is _____?

(8)

93.

Term No.:	1	2	3	4	5	6	7	8
Series :	1	2	4	8	16	32	64	___
	↗	↗	↗	↗	↗	↗		
Σ :	1	3	7	15	31	63	___	___

When n = 5, $\overset{n}{\Sigma} = 31$.
This can also be obtained by subtracting 1 from the next _____ _____ which is 32.

(following T)

94.

Term No.:	1	2	3	4	5	6	7	8
Series :	1	2	4	8	16	32	64	___
Σ :	1 ↗	3 ↗	7 ↗	15 ↗	31 ↗	63 ↗	___	___

When n = 6, $\overset{n}{\Sigma}$ = 63.
Notice that this same relationship is true again. If n = 6, how do you write "the next following T?"

(T_{n+1})

95.

Term No.:	1	2	3	4	5	6	7	8
Series :	1	2	4	8	16	32	64	___
Σ :	1 ↗	3 ↗	7 ↗	15 ↗	31 ↗	63 ↗	___	___

When n = 5, $\overset{n}{\Sigma}$ = 31.
When n = 6, $\overset{n}{\Sigma}$ = 63.
Are you ready to state a general rule: Try it in symbols, letting T_{n+1} mean the term-value of the n + 1st term.
"$\overset{n}{\Sigma}$ ______?"

($\overset{n}{\Sigma} = T_{n+1} - 1$)

96.

Term No.:	1	2	3	4	5	6	7	8	9
Series :	1	2	4	8	16	32	64	___	___
Σ :	1 ↗	3 ↗	7 ↗	15 ↗	31 ↗	63 ↗	___	___	___

You have stated the rule for this series, that $\overset{n}{\Sigma} = T_{n+1} - 1$.
Suppose that the value of the n + 1st term is 1024. What is $\overset{n}{\Sigma}$? ____

(1023)

97.

Term No.:	1	2	3	4	5	6	7	8	9
Series :	1	2	4	8	16	32	64	___	___
Σ :	1 ↗	3 ↗	7 ↗	15 ↗	31 ↗	63 ↗	___	___	___

If the value of Term_n is 256, what is $\overset{n}{\Sigma}$? _____

WATCH IT NOW!

(511)

98. Term No.: 1 2 3 4 5 6 7 8

Series : 1 2 4 8 16 32 64 __

Σ : 1 ↗ 3 ↗ 7 ↗ 15 ↗ 31 ↗ 63 ↗ __ __

If n = 7, what is $\overset{n}{\Sigma}$? ____

(127)

EXHIBIT C

Frames 91 to 100 in the R and E (Rule and Example) Program in which the rule $\overset{n}{\Sigma} = T_{n+1} - 1$ is taught for summing the number series 1, 2, 4, 8, 16, . . .

91.

Term No.:	1	2	3	4	5	6	7	8
Series :	1	2	4	8	16	32	64	___
		↗	↗	↗	↗	↗	↗	
Σ :	1	3	7	15	31	63	___	___

Now we want to find a rule for the sum of this series, in terms of either term number (n) or term-value (T), as indicated by the arrows. The rule is $\Sigma = \text{Term}_{n+1} - 1$ or $\Sigma = T_{n+1} - 1$.

Write the rule on your answer sheet.

($\Sigma = T_{n+1} - 1$)

92.

Term No.:	1	2	3	4	5	6	7	8
Series :	1	2	4	8	16	32	64	___
		↗	↗	↗	↗	↗	↗	
Σ :	1	3	7	15	31	63	___	___

Rule: $\Sigma = T_{n+1} - 1$

In using this rule, you have to locate T_{n+1}, that is, the n + 1st term.
When n = 3 (top row, T (middle row) is 4.
When n = 3, T_{n+1} is 8
When n = 4, T_{n+1} is _____

(16)

93.

Term No.:	1	2	3	4	5	6	7	8
Series :	1	2	4	8	16	32	64	___
		↗	↗	↗	↗	↗	↗	
Σ :	1	3	7	15	31	63	___	___

Rule: $\overset{n}{\Sigma} = T_{n+1} - 1$

When n is 4, T_{n+1} is 16

When n is 5, T_{n+1} is _____?

(32)

94.

Term No.:	1	2	3	4	5	6	7	8
Series :	1	2	4	8	16	32	64	__
Σ :	1 ↗	3 ↗	7 ↗	15 ↗	31 ↗	63 ↗	__	__

Rule: $\overset{n}{\Sigma} = T_{n+1} - 1$

When n is 6, T_{n+1} is 32.

When n is 6, T_{n+1} is _____.

(64)

95.

Term No.:	1	2	3	4	5	6	7	8
Series :	1	2	4	8	16	32	64	__
Σ :	1 ↗	3 ↗	7 ↗	15 ↗	31 ↗	63 ↗	__	__

Rule: $\overset{n}{\Sigma} = T_{n+1} - 1$

When n is 3, what is $\overset{n}{\Sigma}$ _____?

(7)

96.

Term No.:	1	2	3	4	5	6	7	8
Series :	1	2	4	8	16	32	__	__
Σ :	1 ↗	3 ↗	7 ↗	15 ↗	__	__	__	__

Rule: $\overset{n}{\Sigma} = T_{n+1} - 1$

When n is 5, what is $\overset{n}{\Sigma}$ _____?

(31)

97.

Term No.:	1	2	3	4	5	6	7	8
Series :	1	2	4	8	16	32	64	__
Σ :	1 ↗	3 ↗	7 ↗	15 ↗	31 ↗	__	__	__

Rule $\overset{n}{\Sigma} = T_{n+1} - 1$

When n is 5, $\overset{n}{\Sigma}$ is obtained by subtracting 1 from T_{n+1}, which is 32. So, 32 - 1 = 31. When n = 6, what is $\overset{n}{\Sigma}$ _____?

(63)

98. Term No.: 1 2 3 4 5 6 7 8

Series : 1 2 4 8 16 32 64 ___

↗ ↗ ↗ ↗ ↗ ↗

Σ : 1 3 7 15 31 63 ___ ___

Rule: $\overset{n}{\Sigma} = T_{n+1} - 1$

When n = 8, what is $\overset{n}{\Sigma}$ _____?

(255)

---

99. Term No.: 1 2 3 4 5 6 7 8 9

Series : 1 2 4 8 16 32 64 ___ ___

↗ ↗ ↗ ↗ ↗ ↗

Σ : 1 3 7 15 31 63 ___ ___ ___

Rule: $\overset{n}{\Sigma} = T_{n+1} - 1$

Now suppose we ask the question the other way around.

If $\overset{n}{\Sigma}$ is 31, what is T_{n+1} _____? What is T _____?

($T_{n+1} = 32$)

(T = 16)

---

100. Term No.: 1 2 3 4 5 6 7 8

Series : 1 2 4 8 16 32 64 ___

↗ ↗ ↗ ↗ ↗ ↗

Σ : 1 3 7 15 31 63 ___ ___

Rule: $\overset{n}{\Sigma} = T_{n+1} - 1$

If $\overset{n}{\Sigma} = 127$, what is n _____?

(7)

EXHIBIT D

Frames 36 to 40 in the revised GD (Guided Discovery) Program in which the rule $\overset{n}{\Sigma} = T_{n+1} - 1$ is taught for summing the number series 1, 2, 4, 8, 16, . . .

Answers

Subtract one

Subtract one

Subtract one

The sum of the first n terms in the series is equal to the term value of the n + 1th term minus one.

31

D36.

Term No. (n):	1	2	3	4	5	6	7	8
Series (T):	1	2	4	8	16	32	64	___
		↗	↗	↗	↗	↗		
(Σ):	1	3	7	15	31	___	___	___

Here is the same series with connecting arrows to give you a hint as to the rule for summing the series.

When n = 2, $\overset{n}{\Sigma} = 3$ (Think: What do you do to get 3 from the number the arrow points to?)

When n = 3, $\overset{n}{\Sigma} = 7$ (Think: What do you do to 8 to get 7?)

When n = 4, $\overset{n}{\Sigma} = 15$ (Think: What do you do to 16 to get 15?)

Complete the following rule for finding the sum of any number of terms in this series.

$\overset{n}{\Sigma}$ (the sum of the first n terms in this series) is equal to _____.

Use the rule for figuring the answer to this problem.

When n = 5, $\overset{n}{\Sigma} =$ _____

Answers

127

$T_{n+1} - 1$

The sum of the first n terms in the series is equal to the term value of the n + 1th term minus 1.

or

one less than the value of the n + 1th term.

D37. (n): 1 2 3 4 5 6 7 8

(T): 1 2 4 8 15 32 64 ___

(Σ): 1 ↗ 3 ↗ 7 ↗ 15 ↗ 31 ↗ ___ ___ ___

When n = 5, $\overset{n}{\Sigma} = 31$

When n = 6, $\overset{n}{\Sigma} = 63$

When n = 7, $\overset{n}{\Sigma} =$ ___

Are you ready to state a general rule using symbols? Let T_{n+1} mean the term value of the n + 1th term and write a rule in symbols for summing n terms in this number series.

$\overset{n}{\Sigma} =$ ______

Now try stating the rule again in your own words. Complete the rule. "The sum of the first n terms in the series is equal to ____."

15

1023

255

D38. (n): 1 2 3 4 5 6 7 8

(T): 1 2 4 8 16 32 64 ___

(Σ): 1 ↗ 3 ↗ 7 ↗ 15 ↗ 31 ↗ ___ ___ ___

You have stated the rule for this series that $\overset{n}{\Sigma} = T_{n+1} - 1$

When $T_{n+1} = 16$, $\overset{n}{\Sigma} =$ ______

When $T_{n+1} = 1024$, $\overset{n}{\Sigma} =$ ______

When $T_n = 128$, $\overset{n}{\Sigma} =$ ______ (This one is tricky.)

Answers

The sum of the first n terms in the series is equal to one less than the $n + 1^{th}$ term of the series.

$T_{n+1} - 1$

63

D39.

(n):	1	2	3	4	5	6	7	8
(T):	1	2	4	8	16	32	64	___
		↗	↗	↗	↗	↗		
(Σ):	1	3	7	15	31	___	___	___

Rule: in symbols: $\overset{n}{\Sigma} = T_{n+1} - 1$

in words: The sum of the first n terms in the series is equal to one less than the term value of the $n + 1^{th}$ term.

Copy the rule here:

in words: ______________________________

in symbols: ______________________________

When n = 6, $\overset{n}{\Sigma}$ = ________

$T_{n+1} - 1$

The sum of the first n terms in the series is equal to one less than the value of the $n + 1^{th}$ term.

127

255

511

511

2047

D40.

(n):	1	2	3	4	5	6	7	8
(T):	1	2	4	8	16	32	64	___
		↗	↗	↗	↗	↗		
(Σ):	1	3	7	15	31	___	___	___

Rule: The rule for getting the sum of this series is:

in symbols: $\overset{n}{\Sigma}$ = _____

in words: The sum of the first n terms in the series is equal to ____________________

Examples:

When n = 7, $\overset{n}{\Sigma}$ = _____

When n = 8, $\overset{n}{\Sigma}$ = _____

When n = 9, $\overset{n}{\Sigma}$ = _____

When $T_n = 256$, $\overset{n}{\Sigma}$ = _____

When $T_{n+1} = 2048$, $\overset{n}{\Sigma}$ = _____

EXHIBIT E

Frames 36 to 40 in the revised R and E (Rule and Example or Expository) Program in which the rule $\overset{n}{\Sigma} = T_{n+1} - 1$ is taught for summing the number series 1, 2, 4, 8, 16, . . .

Answers

$T_{n+1} - 1$

The sum of the first n terms in the series is equal to one less than the term value of the n + 1th term in the series.

$\overset{n}{\Sigma} = 64 - 1$

$\overset{n}{\Sigma} = 63$

E36.

(n):	Term No.:	1	2	3	4	5	6	7	8
(T):	Series :	1	2	4	8	16	32	64	
(Σ)	:	1 ↗	3 ↗	7 ↗	15 ↗	31 ↗			

Here is the same series with connecting arrows to clarify the rule for summing the series.

Rule: The rule for getting the sum of this series is: in symbols: $\overset{n}{\Sigma} = T_{n+1} - 1$; in words: the sum of the first n terms in the series is equal to one less than the term value of the n + 1th term in the series.

Example: This is how the rule works. Suppose we want the sum of the first 5 terms in the series. Then n = 5 and $T_{n+1} = 32$. Thus, for n = 5, in this series, $\overset{n}{\Sigma} = 32 - 1$ or simplifying the answer $\overset{n}{\Sigma} = 31$.

Rule: Copy the rule: in symbols: $\overset{n}{\Sigma} =$ _____
in words: "the sum of _____"

Example: Use the formula for working out your answer to this problem. When n = 6, $\overset{n}{\Sigma} =$ _____ or simplifying the answer $\overset{n}{\Sigma} =$ _____

Answers

The sum of the first n terms in the series is equal to one less than the term value of the n + 1th term.

31

E37.

(n):	Term No.:	1	2	3	4	5	6	7	8
(T):	Series :	1	2	4	8	16	32	64	
(Σ)	:	1 ↗	3 ↗	7 ↗	15 ↗	31 ↗	___	___	

When n = 2, $\overset{n}{\Sigma} = 3$

When n = 3, $\overset{n}{\Sigma} = 7$

When n = 4, $\overset{n}{\Sigma} = 15$

Complete the rule: $\overset{n}{\Sigma}$ (the sum of the first n terms in this series) is equal to ________

____

Use the rule for figuring the answer to this problem.

When n = 5, $\overset{n}{\Sigma} =$ ____________________

$T_{n+1} - 1$

The sum of the first n terms in the series is equal to one less than the value of the n + 1th term.

127

E38.

(n):	1	2	3	4	5	6	7	8
(T):	1	2	4	8	16	32	64	
(Σ):	1 ↗	3 ↗	7 ↗	15 ↗	31 ↗	___	___	

When n = 5, $\overset{n}{\Sigma} = 31$

When n = 6, $\overset{n}{\Sigma} = 63$

When n = 7, $\overset{n}{\Sigma} =$ ___

State the rule for summing the number series in symbols: $\overset{n}{\Sigma} =$ ____________________

in words: "The sum of the first n terms in the series is equal to ____________________

15

1023

255

E39.

(n):	1	2	3	4	5	6	7	8
(T):	1	2	4	8	16	32	64	___
(Σ):	1 ↗	3 ↗	7 ↗	15 ↗	31 ↗	___	___	___

You have stated the rule for this series that $\overset{n}{\Sigma} = T_{n+1} - 1$

When $T_{n+1} = 16$, $\overset{n}{\Sigma} =$ _____

When $T_{n+1} = 1024$, $\overset{n}{\Sigma} =$ _____

When $T_n = 128$, $\overset{n}{\Sigma} =$ _____ (This one is tricky.)

Answers

$\overset{n}{\Sigma} = T_{n+1} - 1$

The sum of the first n terms in the series is equal to one less than the term value of the n + 1th term in the series. (NOTE: The words need not be exactly as above as long as the meaning is correct.)

127

255

511

511

2047

E40. (n): 1 2 3 4 5 6 7 8 9

(T): 1 2 4 8 16 32 64 128 256

(Σ): 1 ↗ 3 ↗ 7 ↗ 15 ↗ 31 ↗ 63 ↗

Rule: The rule for getting the sum of this series is

in symbols: $\overset{n}{\Sigma} =$

in words: the sum of the first n terms in the series is equal to _____

Examples: Use the rule for working out your answers.

When n = 7, $\overset{n}{\Sigma} =$ _____

When n = 8, $\overset{n}{\Sigma} =$ _____

When n = 9, $\overset{n}{\Sigma} =$ _____

When $T_n = 256$, $\overset{n}{\Sigma} =$ _____

When $T_{n+1} = 2048$, $\overset{n}{\Sigma} =$ _____

References

Ausubel, D. P., "Some psychological aspects of the structure of knowledge," in Stanley Elam (Ed.) Education and the Structure of Knowledge, Chicago: Rand McNally and Company, 1964, 220-264.

Ausubel, D. P., The Psychology of Meaningful Verbal Learning. Greene and Stratton, New York, 1963.

Della-Piana, G. M., "Searching orientation and concept learning," J. educ. Psychol., 48, 1957, 245-253.

Evans, J. L., Glaser, R., and Homme, L. E., "The RULEG system for the construction of programmed verbal learning sequences," Department of Psychology, University of Pittsburgh, August, 1960, Mimeo.

Gagné, R. M., "The acquisition of knowledge," Psychol. Rev., 69, 1962, 355-365.

Gagné, R. M. and Brown, L. T., "Some factors in the programing of conceptual learning," J. exp. Psychol., 62, 1961, 313-321.

Glaser, R. (Ed.), Training Research and Education. University of Pittsburgh Press, 1962.

Schwab, J. J., "Problems, topics and issues," in Stanley Elam (Ed.) Education and the Structure of Knowledge, Fifth Annual Phi Delta Kappa Symposium on Educational Research, Chicago: Rand McNally and Company, 1964 (a), 4-43.

Schwab, J. J., "Structure of the disciplines: meanings and significances, and structure of the natural sciences," in G. W. Ford and Laurence Pugno (Eds.), The Structure of Knowledge and the Curriculum, Chicago: Rand McNally and Company, 1964 (b), 6-49.

Stolurow, L. M., Teaching by Machine. Cooperative Research Monograph, No. 6, OE-34010, U. S. Department of Health, Education, and Welfare, U. S. Government Printing Office, Washington, 1961.

Woodruff, A. D., "A cognitive theory of instruction," Lawrence Siegel (Ed.) Contemporary Theories of Instruction, San Francisco: Chandler Publishing Company (in press).

CHAPTER IV

DISCOVERY VS. EXPOSITORY SEQUENCING IN A PROGRAMED UNIT ON SUMMING NUMBER SERIES*

Garth M. Eldredge

The experiments reported in this chapter are a partial replication and extension of the Gagné-Brown study (1961) utilizing a programed sequence on summing number series. The first section presents a brief report of a replication of the Gagné-Brown study. The second section reports the major study utilizing a revision of the Gagné-Brown programs along the lines spelled out in the previous chapter. The final section reports some exploratory work on problem solving processes of subjects in the different treatments.

STUDY ONE: A BRIEF REPORT OF A REPLICATION

This study utilized the Guided Discovery (GD) and Rule and Example (R and E) programs of the Gagné-Brown study in a modified replication of their work. This study deviates from the Gagné-Brown study in two ways. They used only boys; we used boys and girls and combined them for analysis after finding that there were no sex differences in our sample in mathematical ability. Also, the procedure for administering the criterion test and the scoring procedure differed a little from their procedure and we devised "hints" for the problem solving tests since their hints were not available.

---

*This is an edited and extended version of a doctoral dissertation submitted by Garth M. Eldredge to the Graduate School University of Utah (June, 1965) in partial fulfillment of the requirements for the doctor of philosophy degree in Educational Psychology. Study One was conducted by Marjean Ballard and Study Three by Prudence Redford.

Hypothesis

The major hypothesis of the replication is that there are differences between the GD and RE treatments on criterion problems, favoring the GD groups. In other words the results are hypothesized to be consistent with the Gagné-Brown findings. As indicated in the previous chapter, this cannot be considered an interpretable test of differences between "discovery" and "exposition" methods since the programs used by Gagné and Brown have been demonstrated to contain uncontrolled variables favoring the GD treatment that are not unique to the treatment.

Procedures

Subjects.

The sample was composed of two ninth grade algebra classes taught by the same instructor. Students were selected for these classes on the basis of math ability and randomly assigned to classes. Each class has 13 boys and 13 girls. Thus, one class was used for the GD treatment and the other for the RE treatment.

Materials.

See the original Gagné-Brown study for a description of materials plus the sample given in the previous chapter of items from the Gagné-Brown "summing number series" programs for GD and RE treatments.

General Procedures.

Testing took place on three consecutive days requiring 50 minutes per section on the first day and 90 minutes per section on the following days.

The first day students were given six reference tests of cognitive factors published by the Educational Testing Service, Princeton, New Jersey. The tests included were: 1. Math aptitude (story problems, 2. Vocabulary, 3. Addition, 4. Subtraction, 5. Multiplication and division, 6. Object-number test (memorization of numbers corresponding to words in a list).

The second day subjects were given an introductory program designed to teach concepts relating to summing number series. The program consisted of 92 cards connected on rings. Time on the program was recorded. When a student made an error, he turned back to the item, crossed out his response and wrote the correct one beneath it. Upon completion of the introductory program, subjects were given the treatment assigned.

On the third day subjects went through the same learning program once again but did not repeat the introductory program. Time on program was recorded.

After completing their respective treatments, the subjects were given a short break. Then they received instructions for the criterion problem solving test which consisted of four problems presented one at a time. The subjects were asked to try to solve the first problem by using the proper formula for summing the numbers in the series. If at the end of five minutes the subject did not have the answer, he was given a hint and allowed three more minutes. If at the end of the additional three minutes his formula was still incomplete, he was given a second hint and two more minutes to complete his formula. Time was called at the end of ten minutes. The same procedure was followed for the remaining three problems.

Scoring.

On the criterion problem solving test a record was kept for each subject on: Time to complete the formula, number of hints used and correct or incorrect solution.

Scores were weighted as follows: three points were given for a correct formula completed in less than five minutes; i.e., without hints. Two points were given for a correct answer requiring one hint and less than eight minutes. One point was given for a correct answer requiring two hints and from eight to ten minutes. A score of zero was assigned for an incorrect or incomplete formula at the end of ten minutes.

Results

Because this replication study was incorporated into another study in which we tried out preliminary versions of a revised program, we will pull out from that study those results relevant to the replication without detailed reporting here. A major purpose of this study was to give us first hand acquaintance with the materials prior to our revision of the programs and to try out some preliminary revisions of the Gagné-Brown materials.

The major hypothesis was partially confirmed. The GD treatment was significantly better (.01 level) on problem one and the direction of differences favored the GD treatment on problem two and total problems (but not significantly). The RE treatment was favored (directionally, but not significantly) on problems three and four. Actually Gagné and Brown looked only at total problems. On that basis the present study lends some slight support to their findings.

Applying the W^2 formula (Hays, 1963) to the Gagné-Brown data and data from the present study we get a more discouraging picture. In the present study although t-tests showed significance favoring the GD treatment on one problem, the actual strength of association (per cent of variance in obtained scores accounted for by treatments) never exceeded 13 per cent. In the original Gagné-Brown study, applying the same formula, it was found that 82 per cent of the variance on their total problems weighted hint score could be accounted for by treatment differences. Thus, from the viewpoint of "strength of association," this study lends little support to their findings. In spite of minor procedural and scoring differences, such an outcome lends little confidence in the generality of their findings.

Another finding of the replication was that time to complete the program was less in the present than in the Gagné-Brown study but that our findings supported their report of significantly shorter time for the RE treatment.

Finally, an exploratory probe for additional hypotheses resulted in the unusual finding that mathematics aptitude was significantly positively related to performance on total problems for the GD group ($r = .42$) and significantly negatively related to total problems for the RE group ($r = -.33$). Such a finding suggests further exploration of a treatment-aptitude interaction. It is possible that the negative correlation of aptitude and performance for the RE group was due to negative attitudes developed by students of high math aptitude in the RE treatment.

The major import of the replication study for the study which follows is that it forced us to be very clear about procedures so that the study could be replicated. This we have done as will be clear by the detailed statement of procedures and the publication of the complete programs in a separate volume of the final report.

STUDY TWO: THE REVISED GAGNE-BROWN PROGRAMS

Hypotheses

For a description of the guided discovery and expository programs used in this study see the previous chapter. For a copy of the complete program see Part II of the report published as a separate volume.

1. On an immediate transfer test of problem solving, there are no differences in performance between the two treatments. The rationale for this is based on the fact that both have practiced and produced actively the kinds of responses which are called for on the criterion test.

2. There are no differences between treatments on a delayed transfer test (after 4 weeks). It is true that one might expect differences between treatments as a result of the development of a searching set in the guided discovery treatment and the development of "affective" responses enhancing the guided discovery treatment by encouraging "attention" and "rehearsing." However, it is expected here that due to the adequate treatment controls and the short time period of instruction such advantages could not adequately develop.

3. High ability subjects perform better on criterion tests of transfer and retention after having been subjected to the Expository (R and E) treatment and low ability subjects perform better on tests of transfer and retention after having been subjected to the Guided Discovery (GD) treatment. The rationale for the hypothesis concerning ability and treatment interaction effects stems from the notion that high ability subjects can think faster and take more short cuts in problem solving than low ability subjects. They do not require the same amount of manipulation of the materials to achieve mastery. Although the programs were designed to equate the amount of computation, the searching orientation

may be an added "chore" to the higher ability subjects.

4. There are no sex differences on the criterion tasks. There is no evidence in the literature to indicate that males differ from females in ability to handle complex mental materials when they are at the same general ability levels.

5. On an immediate test of transfer, there are no differences between the subjects from one school as compared with the other school when adjustments are made for ability differences. It is predicted that there are ability level differences, however, favoring the upper socio-economic group.

6. Subjects who are given an opportunity to explain verbally the solutions to the criterion problems directly following the test of immediate transfer improve scores significantly more than subjects who are not given this opportunity.

Procedures

The design of the study is described in this section under six headings: 1. the subjects; 2. the material used in the study; 3. the general procedure and instructions to subjects; 4. the scoring of the criterion measures; 5. the reliability of the instruments; 6. the experimental design.

Subjects.

Subjects were ninth grade algebra students in two junior high schools in the Granite School District, Salt Lake City, Utah. All ninth grade algebra students in the two schools were used in the study.

One school (School 1) was from a low socio-economic area of the district and the other (School 2) was from an upper socio-economic area based on family income and years of education as described by Richardson and Pawar (1964). There were twenty-eight subjects (8 boys and 20 girls) from School 1 and sixty-nine subjects (35 boys and 34 girls) from School 2 who completed all parts of the program. A table of random numbers was used to eliminate from the analysis the data of one male subject at School 2. This allowed 4 boys and 10 girls for each treatment (rule and example and guided discovery) from School 1, and 17 boys and 17 girls for each treatment from School 2. Data obtained from all 96 subjects were used in the analysis.

Material Used in the Study.

Materials used in the study were the Verbal Reasoning and Numerical Ability sub-tests of the Differential Aptitude Test Battery (DAT), Form L. These were used to give the experimenter some measure of general ability to check on the effectiveness of random assignment to treatments and provide a basis for statistical control if necessary.

The DAT has been subjected to much research. The manual reports reliability data for the VR + NA scores as .88 for ninth grade subjects. Unknown to the experimenter at the time of the study, the subjects had been given the DAT some three weeks prior to the experiment. The two sets of scores, however, served as a test-retest reliability check for the DAT with a three-week interval. The reliability coefficient obtained in this manner is almost identical to that given in the manual (.89).

The DAT also correlates highly with other tests which have been demonstrated to measure overall ability. The reported correlations of the VA + NA scores with the Wechsler Adult Intelligence Scale range from .74 to .79 which indicates that these DAT sub-test scores can be assumed to be a measure of overall ability.

Subjects from School 1 had a mean score at the 70th percentile on a table of national 9th grade norms and those from School 2 had a mean score at the 85th percentile on the same norms. Hence, subjects were, in general, a very capable group.

Also used in the study were a series of learning programs adapted from those used by Gagné and Brown (1961). An introductory learning program was designed to teach the terminology to be used in later learning tasks. The complete program is described in the previous chapter and presented in the Appendices published as a separate volume. The programs are revisions of those used by Gagné and Brown (1961). The R and E and introductory programs had been revised once before for an earlier study (Ballard, 1964). After a second revision for these programs and a first revision for the GD program, they were presented to a pilot group of twelve 9th grade students from the Stewart Laboratory School of the University of Utah. The pilot group was used to determine error rate and gather other data useful to guiding a revision. A few minor revisions were made based on results obtained from the pilot group before the final programs were printed.

The programs consisted of a 34 frame introductory program suitable for both treatments and separate 35 frame GD and R and E learning programs. Each frame required multiple responses, however, yielding 257 responses for the introductory program and GD treatment and 255 responses for the introductory program and R and E treatment. Because subjects were required to go through the special treatment program twice, GD subjects made a total of 367 responses and R and E subjects made a total of 363 responses when one includes the second time through the treatment program.

Materials were printed on 5" x 7" cards with the correct responses to the frames printed on the back of the cards. Cards were held together with rings and subjects merely flipped the cards over as each frame was completed. If a subject in either treatment made an error, he was instructed to turn back and reread the item, and then correct the response at the bottom of the card.

A series of transfer problems was constructed to test the learning of the rules and formulas by the subjects. Problems were printed on 5" x 7" cards. For each problem two hints were devised to aid in finding a short-cut rule and formula for summing the new series. (These hints were used only if solutions were not reached within certain time limits.) All cards were separated by blank sheets of paper. Each problem, together with the hints accompanying that problem were sealed together with a piece of cellophane tape. Under each tape was an identifying mark so as to eliminate any confusion as to which problem was being administered at any one time. The series of problems were held together with metal rings inserted into holes punched in the cards. The problems and hints are also found in the Appendices published as a separate volume.

The four problems which were used in the learning task itself were reprinted on 5" x 7" cards to be presented after four weeks as a retention measure. There were no hints provided for these test problems which also are published in the separate volume of Appendices.

All learning tasks and the test of immediate transfer were administered during the first two weeks of November, 1964. The experiment lasted for three days with each group of subjects. Test of retention and delayed transfer were administered four weeks after the initial learning took place.

General Procedure and Instructions to Subjects.

Subjects were first given the pre-tests which consisted of the verbal reasoning and numerical ability sub-tests of the Differential Aptitude Tests. Instructions were as follows:

Good morning! We are from the University of Utah and are to conduct an experiment in programed learning. We are trying to find out something about how students learn. The results of this experiment may help us design more helpful and interesting teaching materials. Please follow very carefully the instructions which will be given.

This experiment will continue for three days. You will be here for approximately two periods today and approximately three periods tomorrow and three periods on Wednesday. We appreciate your cooperation.

Today we have a series of exercises for you to do. Will you listen carefully to the instructions.

At this point, the DAT was administered.

On the second day of the experiment, subjects were given the following instructions:

Good morning! Today each of you will be given two sets of cards which have been designed to teach you something about number series. You have just received the first set of cards. Write your name in the space provided on the top card now.

(Pause)

When you begin to work on the cards, read each card, and write your answer in the space provided. Then flip the card over and read the answer on the back of the card. If your answer and the answer on the back of the card are the same, go on to the next card. If not, turn the card back over and reread the item, then write the correct answer at the bottom of the card. Do not erase your first answer. This is an experiment in a new way of teaching and we need to know just where students are making mistakes so that we can improve the material. Try to be as accurate as possible, but do not worry about the mistakes you make on the program. The number of mistakes is not important. It is important that you learn the material, however, since you will be questioned about it at a later time.

After completing one card and making the necessary corrections, flip the card all the way over and under the others like this . . . (demonstrate) . . . before going on to the next one.

When you have finished all the cards in the first set, record the time that is written on the blackboard, and turn the cards in at the front of the room. You will be handed a second set of cards for which you are to repeat the same procedures.

(Proctors check to see that time is recorded, first set turned in, beginning time recorded on second set, student not erasing but is "correcting," end time written and no peeking at answers before writing own answer.)

Be certain that you record the time that you end the first set of cards and the time that you begin and end the second set of cards.

If you have any questions as you proceed, raise your hand for some help.

Are there any questions before we begin?

(Pause)

Write in the proper space the beginning time which is _____ and begin working.

Subjects were given the introductory program and allowed to work at their own speed. As subjects neared the end of the program, a proctor began writing the time on the blackboard and did so every half minute until all subjects had completed both the introductory and the learning programs. Proctors were in the room (one for each six to seven subjects) throughout the entire time that subjects were involved in the learning programs and were trained to answer specific questions which might arise in the learning process.

Subjects were assigned to one of the two treatments on the following basis. Subjects were seated at a series of tables. Near each table was placed enough learning materials for all subjects at that table. As a male subject finished the introductory program, he was randomly assigned to one of the two treatments. The next male subject to finish was assigned to the second treatment and subjects were thus alternately assigned to a learning program as the introductory program was completed. The same procedure was followed for the female subjects. This process was used to assure the experimenter that there would be no advantage to either treatment on the basis of time involved in learning the introductory program.

On the third day of the study, subjects were again given the same learning task as on the second day to aid more individuals to achieve some competence in working with number series. The introductory program was not repeated.

Instructions were as follows:

> Each of you has been given a set of cards. Write your name in the proper space now.
>
> (Pause)
>
> These cards are simply another copy of the second set of cards on which you worked yesterday. The instructions are the same.
>
> When you begin to work on the cards, read each card, and write your answer in the space provided. Then flip the card over and read the answer on the back of the card. If your answer and the answer on the back of the card are the same, go on to the next card. If not, turn the card back over and reread the item; then write the correct answer at the bottom of the card. Do not erase your first answer. This is an experiment in programed learning and we need to know where students make mistakes so that we can improve the program. The number of mistakes you make is not important. It is important that you learn the material, however, since you will be questioned about it at a later time.
>
> After completing one card and making the necessary corrections, flip the card all the way over and under the others like this . . . (demonstrate) . . . before going on to the next card.
>
> When you have finished all the cards, record the time that is written on the blackboard, turn your cards in at the front of the room and wait for further instructions.
>
> Are there any questions?
>
> Record the time which is _____ and begin.

As soon as all subjects completed the second learning task, the materials were gathered and subjects were given the series of immediate transfer problems. The problems consisted of number series tasks for which subjects were to find a rule and formula for summing the series in a short-cut manner. The transfer test consisted of four problems which required the subject to find formulas for summing number series which were different from those used in the learning task. Subjects were instructed as follows:

> You will now have a chance to work some problems. The problems are on the cards you have just received. Write your name in the proper space on the top card now. *Do not turn any card over until told to do so*.
>
> In each of these problems you will be given a new number series for which you are to find a formula using term number (n) or term value (T_n) or both for summing the first n terms of the series. You may make notes on the card if you wish. Remember, your task is to write a rule or formula for summing the series.
>
> Here is an example:
>
> | | | | | | |
> |---|---|---|---|---|---|
> | n: | 1 | 2 | 3 | 4 | 5 |
> | T: | 0 | 2 | 4 | 6 | 8 |
> | Σ: | 0 | 2 | 6 | 12 | 20 |
>
> Your task is to find a rule for summing the first n terms of the series using term number (n), term value (T_n) or both but not by adding each term value in the series.
>
> the top row represents the term number
>
> the middle row represents the term value
>
> the bottom row represents the sum
>
> Now suppose we wanted to get the sum of the first three numbers in this series, 0, 2, and 4 . . . (underline them on the board) . . . without adding them up separately. Actually, of course, we know that the sum is six . . . (point and circle). But how could we get six without adding up each number separately? In this case n is 3, and 3 times 2 will give us six. That is, the sum for the first n terms is equal to n (n - 1) . . . (point). Let's see if it works with the first four terms: n or (4) times n - 1 or (3) = 12 which *is* the sum of the first four times in this series.
>
> Your task is to work several problems like the one I have just shown you.
>
> You will have approximately five minutes to work each problem. Begin when I say begin and stop when I give the signal. If you finish a problem before time is called, record the time that is written on the blackboard and raise your hand so that one of us can check your work.

> (Proctors check to see if formula derived by student actually provides a solution.)
>
> If your answer is correct, wait for further instructions. If your answer is not correct, you will be told to rework the problem. In that case you will record the final time it takes you to complete the problem.
>
> Remember, you are to write a rule using the term number, term value, or both for summing the first n terms of the series. Work as fast as you can as your score will depend upon the amount of time used as well as the correctness of the solution.
>
> Are there any questions?
>
> Using the point of your pencil, break the seal which holds together the first group of cards (there is a red circle under the seal.)
>
> (Pause)
>
> Turn the page and in the upper right hand corner of the top card record the time which is _____ and begin.
>
> (Proctors check and then hand out face down the anecdotes.)
>
> After 5 minutes
>
> Stop! Write just beneath the time recorded in the upper right hand corner the ending time which is _____. Those who think you have a solution raise your hands and we will check your work.

Subjects were given five minutes in which to write a short-cut formula for summing the numbers in the series, at which time, those who had not reached criterion were given a hint. After three more minutes of work, a second hint was given. Time was called after two more minutes. Total time allowedwas ten minutes per problem. The instructions after five minutes were as follows:

> Those who have found a correct solution may read from the material which you have just received (the anecdotes). Those who have not completed a solution will be given a hint which may help you see a relationship between the term number or the term value or both and the desired sum. When you have a solution, record the time that is on the blackboard and raise your hand so one of us can check your work.

Flip the first card and the blank card over and record in the upper right hand corner the time which is _____ and begin.

After 3 minutes

Stop! Record the time which is _____. Those of you who think you have a solution should raise your hands so we can check your work. Those whose solutions are correct may read from the material which you have received. The remainder will be given a second hint which may help you to see a relationship between the term number or the term value or both and the desired sum.

When you have a solution, raise your hand and record the time that is on the blackboard.

Turn over past the blank page and in the upper right hand corner record the time which is _____ and begin.

After 2 minutes

Stop! Record the time which is _____.

Put down your reading materials.

Subjects who completed the criterion task before the time limit expired were given a series of anecdotes to read. The anecdotes served to provide subjects with a constant intervening variable thus preventing them from working algebra or other math homework during the intervening periods of time.

Proctors were in the room to check the work of the subjects. As soon as a subject believed he had an adequate solution to a problem, he would record the time that was written on the blackboard and raise his hand so that a proctor could check his work. Trained proctors checked the work against a list of possible short-cut solutions. If the subject's solution did not match one on the list, the proctor would test the solution to determine whether or not it was acceptable. If the solution was acceptable, subjects were allowed to read from the selection of anecdotes. If the solution was incorrect, subjects were required to rework the problem.

After the first problem was completed, instructions were given as follows:

> You will now be given more problems of the same type. Your task is to write a rule for summing the first n terms of the series using the term number or term value or both.
>
> Try to find the answer as quickly as possible as your score will depend upon the amount of time and the number of hints used as well as the correctness of your solution.
>
> Raise your hand when you have a correct formula and record the time which is written on the blackboard. One of us will check your solution. When your solution is correct, you may continue reading from the material which you have received. Do not go on to another problem until told to do so.
>
> (Pause)
>
> Break the seal (the one with the blue circle) and turn to problem 2. Record the upper right hand corner the time which is _____ and begin.

The remaining problems were administered in the same manner as problem 1.

Scoring.

The score was based on the total number of minutes required to solve the problem, plus a three minute penalty if hint one was used, plus a two minute penalty if hint two was used, plus a one minute penalty if no solution was reached, plus a two minute penalty if the solution was right as to process but used the wrong terminology. Maximum score for any problem was 18, indicating that no solution was reached. (Those who could not write a solution in the allotted time were also penalized as though they had used the wrong terminology.) Thus, the lower the score the better the problem solving performance.

The scoring procedure is almost identical with that used by Gagné and Brown (1961) with the exception of the time penalty added for not having learned the terminology (i.e., algebraic notation) which was used in the learning program.

Four weeks following the initial learning, subjects were administered the same four transfer problems as the measure of delayed transfer. Subjects had no prior knowledge that a second testing session would be required. The delayed transfer test problems used the same hints as previously used and were scored in the same manner.

The same number series problems taught in the experimental learning program were also administered at this time as a test of retention. There were no hints supplied for the retention test. Scores on the retention test were based only on the amount of time needed to solve the problem plus a one minute penalty for no solution and a two minute penalty if the solution was right as to process but wrong as to terminology. The time limit for each problem was five minutes. Maximum score possible was eight points per problem for a total of thirty-two which indicated poor performance.

Reliability of Instruments.

One of the first items of concern in any study is the reliability of the criterion measures. Some measure of internal consistency must be made to assess the reliability of the instruments used.

The most common standard formula for measuring a coefficient of internal consistency, the Kuder-Richardson formula, is not applicable for weighted scores. Cronbach (1951) describes a general formula for coefficient alpha (α) (of which the Kuder-Richardson coefficient of equivalence is a special case) which can be used for weighted scores. The general formula for coefficient alpha (α) is shown to be the mean of all split-half coefficients resulting from different splittings of the test. Alpha is therefore an estimate of the correlation between two random samples of items from a universe of items like those in a test and is computed from the following formula:

$$\alpha = \frac{n}{n - 1} \left(1 - \sum_i \frac{Vi}{Vt}\right)$$

Here n is the number of items in the test, Vi is the variance of item scores after weighting and Vt is the variance of test scores.

Because the criterion measures consisted of only four weighted problems each, there was some question as to how reliable they might be.

The alpha computed for the first criterion measure was .56. Alpha was again computed for the same criterion measure when it was used as a test of delayed transfer (after four weeks). The second computed alpha was .55. Hence, though the measures did not exhibit extremely high reliability coefficients, it was assumed that they were adequate for the type of data analysis used in this study.

Experimental Design.

The data was coded in numerical form and punched on IBM cards and processed at the Computer Center at the University of Utah.

Hypotheses were tested by means of simple analysis of covariance design. The covariance design was used to control for ability differences and time spent in working on the learning tasks. It was believed that these variables might differentially affect the results, hence a covariance design was used rather than a simple t-test for comparing the differences between means.

After the data had been processed at the Computer Center, an error was found in one of the covariance problems. As it was apparent that the covariates of interest on this variable were not significantly affecting the means, hand calculations of analysis of variance were made to test the treatment effects.

To test the interaction effects of ability with treatment, a two-way factorial design was used, again utilizing the covariance design to control for time spent on the learning program.

The sample was divided into high and low ability groups based on the pre-tests dichotomized at the mean. The two levels of ability and the two treatments combined to give the 2 x 2 factorial design.

In order to maintain equal numbers of subjects in each of the four cells of the factorial design, a table of random numbers was used to eliminate a total of eight subjects from three of the cells. This left twenty-two subjects per cell for a total of eighty-eight subjects in the analysis.

Basic to analysis of covariance design is the assumption of homogeneity of variance. F tests were made to compare the difference in variances on the dependent variable by the formula from Guilford (1956, p. 224).

$$F = \frac{\text{larger variance}}{\text{smaller variance}}$$

None of the F's computed in this manner reached significance for 47 and 47 degrees of freedom indicating that the assumption was adequately met.

Another assumption basic to analysis of covariance is normality of the population. Scatter diagrams of the samples indicated that they very nearly followed the normal distribution.

Because the samples were not randomly selected (although assignment to treatment was random) and the fact that the groups could not be matched on all possible relevant variables, the covariance design is most applicable to give statistical tests of obtained differences between groups.

Results and Discussion

An examination of the errors made in the programs used in the present study indicate that the programs were moderately effective in shaping up the behavior of the subjects. The error rate was 16 per cent for the total program for each treatment. However, only three subjects failed to achieve any measure of criterion behavior on the delayed retention test. There was no difference between treatments on the total error rate. Error rates are presented in Table 1. (All tables for this study are presented at the end of the chapter.)

Perhaps a more significant internal measure of program effectiveness is error rate on criterion problems. At the end of each learning program, one frame was designed as a criterion frame of what had been learned in the program. There were no significant differences between treatments on number of errors on this measure, indicating that both treatments yielded achievement at about the same level on the criterion tasks. The actual error rate on the criterion frames was 20.8 per cent for the GD treatment and 22.7 per cent for the expository treatment. Because the criterion frame response suffers from being under a high sequence prompt, the subject having just been through an instructional sequence, a delayed retention measure was also used in the study.

The time spent working on the learning programs was tabulated. It was found that the treatments did not differ on this dimension. The times spent working on the programs are shown on Table 2.

Similarly, the treatments did not differ on the ability dimension as measured by the DAT. These data are presented in Table 3. It was assumed that the treatment groups were well equated and comparisons could be made between the methods of instruction.

Test of the Hypotheses.

It was hypothesized that when adjustments were made for ability level differences, and time spent on the learning program that there would be no differences between the R and E and GD treatments on tests of immediate transfer. It was believed that differences in other studies favoring the GD treatments were products of the experimentaldesign and, with a better controlled, more carefully designed experiment, initial differences would disappear. The hypothesis was not supported in the present study. Treatment differences were found which were significant at the .05 level. The findings suggest that a carefully programed guided discovery approach to learning, wherein the guided discovery treatment presumably teaches a searching set, enhances initial learning. The data from this more carefully controlled study supports the findings of Gagné and Brown (1961) wherein the programed GD treatment was superior to the programed R and E treatment. The analysis of covariance data are presented in Table 4.

The trend of evidence reported in the literature is toward no initial differences between R and E and GD treatments but for delayed transfer and retention tests to favor and GD treatment. It was hypothesized that the findings of this study would support that trend for initial differences but not for delayed differences. The initial treatment differences favoring the GD treatment were retained on a test of delayed transfer (after four weeks). Both treatments improved performance slightly. However, when the covariance design controlled for differences between treatments on the initial transfer test, the differences over time disappeared. That there were significant differences in immediate transfer and delayed transfer favoring the GD treatment is not disputed. However, the delayed transfer differences cannot be explained in terms of enhancement of motivation as hypothesized by Kersh (1962) because a look at delayed transfer differences between treatments adjusting for differences in immediate transfer proved non-significant. Thus, delayed transfer treatment differences appear to be due to initial learning rather than differential rehearsing effects. These findings are shown in Tables 5 and 6. Table 5 presents the findings when no controls were made for initial learning differences and Table 6 presents the findings when controls were made for the initial learning differences. There were no significant differences between treatments on a test of retention of the specific formulas taught in the learning program treatment. These findings are shown in Table 7.

Much of the literature concerning discovery learning favors not one method over all subjects, but rather shows an interaction effect. Typically the brighter subjects profit more from direct exposition and the duller subjects profit more from a form of discovery learning. An interaction effect in the direction of the literature was hypothesized, but it was not supported. There were no ability by treatment interaction effects found in the present study. Both high ability and low ability subjects as individual groups performed better after having been subjected to the GD treatment.

The subjects used in this study were considerably above average in ability level. In reality, there were no low ability subjects. The high ability group had mean aptitude scores at the 90th percentile of 9th grade norms and the low ability subjects had mean aptitude scores at the 65th percentile of 9th grade norms. Thus, the restriction in range of ability could account for the lack of interaction effects. It is interesting to note, however, that the trend is away from that found in the literature, as these very capable subjects all performed at a superior level on the GD treatment. These findings are shown in Table 8 and Table 9. There were no differences between treatments on the retention measures although high ability subjects performed at a level superior to those subjects described as low-ability. These findings are presented in Table 10.

As was predicted, there were no significant sex differences found in the analysis of the data. It is assumed that the members of the two sexes can and do learn equally well in the present situation. This is not to imply that such relationships would necessarily hold across all possible subjects, however.

In the typical junior high school, girls tend to do slightly better than boys in tests of academic performance. Boys, however, tend to do slightly better than girls in the mathematics area. Only highly selected students are allowed to register for algebra in the 9th grade, and as a result sex differences should disappear in this subject matter area. The data of the present study support these notions and are presented in Tables 11, 12, and 13.

Subjects for the present study were selected from two schools. School 1 was in a lower middle socio-economic area. School 2 was in an upper socio-economic area. Typically there are ability level differences between students taken from such diverse areas. It was hypothesized that there would be differences in ability levels of the subjects from the two schools, but that when such differences were controlled for in a covariance design, there would be no differences in performance on the criterion tasks. The findings are presented in Tables 14, 15, and 16.

It can be observed from the t-tests in Table 17 that there were indeed differences between subjects from the two schools. The subjects at School 1 had mean DAT scores significantly lower than those obtained at School 2. Even though subjects are selected to register for algebra on the basis of 8th grade scores on the STEP tests, the mean scores on the DAT reflect very real differences between the subjects from the two schools. Subjects from School 1 have a mean score at the 70th percentile on ninth grade norms and subjects from School 2 have a mean score at the 85th percentile on ninth grade norms.

On the tests of immediate transfer the subjects from School 2 performed significantly superior to subjects from School 1. However, the covariance design was used to control for initial ability differences. The resulting findings support the hypothesis that such differences would disappear when controls were made for the contaminating variable (ability level differences).

Following the administration of the initial transfer task, the subjects from School 1 were given an opportunity to explain orally the processes by which they solved the criterion problems. It was hypothesized that this oral explanation would be an added advantage to these subjects and, as a result, they would experience greater gain scores between the initial transfer test and the delayed transfer test than would subjects from School 2.

The latter hypothesis is in contrast to Hendrix's (1961) notions concerning verbalization of a discovered principle. Her contention was that verbalization has an inhibiting effect on retention and transfer of learned principles. Hendrix's hypothesis was not supported in the present study although it could well be that the conditions of verbalization were different than those which Hendrix finds inhibiting. Net change scores were computed for each group of subjects and were found to be significantly different at the .03 level, the change favoring the subjects from School 1, the lower socio-economic area. The verbalization of the principles and processes in the present study appeared to enhance retention and transfer rather than inhibit it. It is, of course, possible that regression effects could account for this result. However, the magnitude of the differences is such as to make this unlikely.

The Retention Test.

The discussion to this point has been generally limited to performance on the tests of initial and delayed transfer. In addition to these two tests, subjects were administered a test of delayed retention (i.e., scores on the problems within the learning program but taken again four weeks later). No treatment differences were found on the retention measure.

Discussion.

In a similar study, Gagné and Brown (1961) found significant differences favoring the GD approach to learning. They stated at the conclusion of their paper: ". . . our results emphasized the importance of 'what is learned' rather than, 'how it is learned' as the crucial factor in learning effectiveness." The interpretation by Gagné and Brown (1961) was criticized by Della-Piana and Eldredge (1964) because of lack of control of some seemingly relevant variables.

The present study controlled for the non-experimental variables apparently uncontrolled in the Gagné and Brown study (1961) and confirmed the results of their study favoring the GD treatment.

The findings of the present study do not generally follow the trend of the literature. Interpretation must be subject to analysis of the programs themselves. It becomes important to ask, "How different are these programs." Because an effort was made to equate the practice effects, which apparently met with success, one wonders just how the programs differed as teaching devices. The fact that significant differences were obtained on tests of immediate and delayed transfer favoring the GD treatment indicates that some selective behaviors were being shaped in the GD program which enhanced transfer to new problem situations more than in the R and E treatment.

The present findings suggest that "how something is taught" does have an impact on learning. This is not meant to imply anything concerning the teacher presentation variable, but rather the actual method of presentation of the subject matter. That is, in the present study, "what is taught" was well controlled between treatments and the sequencing of material was varied to conform to what has loosly been called guided discovery versus direct exposition treatments. The findings that GD is superior to R and E on immediate and delayed transfer is important for learning. Why this method proved superior is not immediately clear. One possibility is that it yields greater motivational value. This interpretation (Kersh, 1962) is not supported by our findings which reveal no significant differences on delayed transfer when immediate transfer is controlled by covariance analysis. That is, the post-treatment rehearsing which is characteristic of the highly motivated, if it did occur, did not differentially affect treatments. Another possiblity is that the GD treatment reinforced greater searching set behavior which in turn facilitated transfer. Fortunately we gathered data to allow a test of that hypothesis. The final section of this chapter reports that study.

Hendrix (1961) has been a proponent of "discovery" methods for many years. However, she has claimed that any advantages of "discovery" are lost if the subject attempts to verbalize the principle involved. Gagné and Smith (1962) did not substantiate her findings. Neither does the present study support Hendrix's notions. Rather, the subjects who were called upon to verbalize the principles and processes enhanced their scores significantly more than those not given the oral problem solving task.

It is possible that verbalization of principles must take place during a critical period of the discovery process to inhibit retention and transfer to new situations. If such is the case, the subjects in the present study may have passed the critical period. This has not been throughly investigated and is offered only as a suggestion for reconciling the differences between the present study and that of Hendrix (1961).

STUDY THREE: PROBLEM SOLVING PROCESSES

Although our project was not originally designed to investigate problem solving processes, it appeared feasible to include an exploratory study of this variable within the major design. We have already reported that subjects in School 1 of the previous study were given an opportunity to explain orally the processes by which they solved the criterion problems in the immediate transfer test. Also, we noted that this intervening activity appeared to have a positive effect on delayed transfer. In the study reported here we report data on treatment differences in reported problem solving processes.

Hypothesis

There was one major exploratory hypothesis to this phase of our research. Following the rationale for hypotheses of the previous study, we again expected no differences in problem solving processes between treatments because of the design of our treatments. The treatments were designed to teach the same behaviors except for the possibility that the different sequencing of the guided discovery treatment might lead to the development of a "searching set" for that group. However, we did not expect differences in sequencing to shape up different problem solving behaviors because the treatments were short and the answers were given soon after an initial "try" to generate a rule for summing number series in the guided discovery treatment.

Procedure

The learning tasks and tests for this study were the same as for the previously reported study except as specified below.

Subjects.

Twenty-eight ninth grade students enrolled in first year algebra served as subjects. These were the students in School 1, the lower middle class socio-economic area referred to in the study reported above.

General Procedure.

Procedures were the same as in the previous study except that following the immediate transfer test private oral problem solving interviews were held with each subject in both treatments in an effort to determine the problem solving processes they used in solving the immediate transfer problems. Using students from a pilot group who had completed the same learning program Della-Piana and Redford tried out various means of stimulating recall and tapping subjects' thought processes. On the basis of this tryout, procedures were developed for interviews.

After the interviews the subjects were dismissed. Five weeks later they returned for the delayed transfer test and delayed retention test as described in the previous study.

Oral Problem Solving Procedures.

Interviewers were trained in the use of the following instructions which were on mimeographed sheets with space for recording subjects' comments on each of the four transfer problems.

> "On problem _____ your answer was _____. Now look at it and think back to how you came to work it out this way. Tell me out loud what you thought and did as you worked out the problem."

Additional probes used as needed by the interviewers were:

> "What was the first thing you did when you came to this problem? Describe in detail what you thought and did. Did you try out any other ways of doing it? Describe them, then what did you do?"

The interviews were set up so that students waiting could not hear students being interviewed. Subjects were asked not to talk to others about their interviews. Each proctor interviewed three to five students and began with problem 4 going backward to 3, 2, and 1 in order to discuss the problem which had most recently received attention first. Interviewers were instructed to write down everything the student said and the next day submit a typed or legibly written copy. Interviewers were not to categorize the behavior. Categories for scoring are discussed below.

Scoring.

Several scoring procedures were tried out using interviewers recordings of oral problem solving processes. For devising scoring procedures four subjects' protocols were used. The subjects used had transfer test scores varying from very high to very low.

Although we finally used only two categories (verification and patterning), the five categories below were devised for trial purposes:

Category	Sample of Behavior
insight	"just saw it" or "immediately I saw that if you subtracted 1/3 . . ."
process flexibility	"tried addition, then subtraction, then multiplication. . ."

Category	Sample of Behavior
verification	"Tried it here (pointing to $n = 3$) and it worked so then I tried it again up here (pointing to $n = 5$) to check it."
patterning	"First I subtracted the sum from $T_n + 1$ all along the series and I noticed a regularity" or "I subtracted Σ from T_n all the way to see if there was a pattern."
problem definition	"Well, I was trying to get this (pointing to value 13 in the Σ row); the idea was to try to get the sum by using these (pointing to n or T rows)."

Several scoring formulas were eliminated because there was a high correlation between scores on transfer tests and scores derived by including the problem solving processes. The scoring scheme finally used was the verification score plus the patterning score. Higher scores indicated better problem solving processes. This scoring procedure yielded least correlation with other criterion measure scores and the interscorer agreement was 100%. Some categories involved behavior which occurred too rarely or consistently yielded low interscorer agreement.

Results and Discussion

Results on immediate and delayed transfer tests, time and error scores, and immediate and delayed retention tests are reported in the previous study. The crucial variable in the present study is the "problem solving process" score. The major finding is that there were no significant treatment differences on oral problem solving using DAT scores and time on task as covariates in an analysis of covariance design. Nor were there significant differences using initial transfer as a covariate. See Tables 19 and 20. There is one study in the literature closely paralleling this phase of our study (Kersh, 1964). However, since Kersh's programs, are not presented, it is difficult to interpret his results.

In our study we did not expect differences because of our assumption that once expository and discovery approaches are equated, except for the crucial experimental variable, no differences in learning would occur. This expectation was supported in our study of problem solving process but not in our study of problem solution accuracy. Since our analysis of problem solution accuracy was based on both schools we went back to the data of the one school used for the exploratory study of problem solving process and found that in that school no differences in problem solution accuracy were found either. Thus, the significant treatment difference in problem solving accuracy seemed entirely due to School 2 subjects. This opens up more questions for us. However, it does tell us that in School 1 (a low socio-economic group) no differences in problem solving accuracy or process was found between treatments. In School 2 problem solving accuracy differences were found but process was not investigated.

TABLE 1

TOTAL ERRORS AND ERROR RATES ON THE LEARNING PROGRAMS

	Treatments					
	R and E		GD		Combined	
	M	%	M	%	M	%
Introductory Program	26.0	17.7	23.5	16.0	24.8	16.9
1st Learning Program	19.4	17.9	21.5	19.5	20.5	18.8
2nd Learning Program	14.5	13.4	15.4	14.0	14.9	13.7
Total	59.9	16.5	69.4	16.5	60.2	16.5

TABLE 2

TIME SPENT ON LEARNING PROGRAMS (MINUTES)

	Treatments					
	R and E		GD		Combined	
	M	s	M	s	M	s
Introductory Learning	55.2	7.9	54.2	9.3	54.7	8.7
1st Learning	64.2	10.7	63.1	11.2	63.7	10.9
2nd Learning	43.4	6.7	43.9	8.1	43.6	7.4
Total	162.8	19.2	161.2	24.1	162.0	21.7

TABLE 3

DAT MEANS AND STANDARD DEVIATIONS

	Treatments					
	R and E		GD		Combined	
	M	s	M	s	M	s
VR	32.8	7.1	33.0	7.1	32.9	7.1
NA	26.1	65.	26.1	5.8	26.1	6.1
VR + NA	58.9	12.2	59.1	11.9	59.0	12.0

TABLE 4

ANALYSIS OF COVARIANCE OF INITIAL TRANSFER TEST FOR 48 R AND E SUBJECTS AND 48 GD SUBJECTS USING DAT (VR + NA) AND TIME ON THE PROGRAM AS THE COVARIATES

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Between Treatments	660.90	1	660.90	5.34
Error (Within)	11391.72	92	123.82	
Total	12052.62	93		

F required for significance at .05 level = 5.29 (two-tailed test).

TABLE 5

ANALYSIS OF VARIANCE OF DELAYED TRANSFER TEST FOR 48 R&E AND 48 GD SUBJECTS DIVIDED INTO HIGH AND LOW ABILITY GROUPS BASED ON PRETEST SCORES

Source of Variation	Sum of Squares	d.f.	Mean Squares	F
Between Treatments	658.8	1	658.8	4.18
Between Abilities	1646.4	1	1646.4	10.44
Interaction	91.6	1	91.6	.58
Error (Within)	14509.1	92	157.7	
Total	16905.9	95		

F required for significance at .05 level = 4.00 (one-tailed test).

TABLE 6

ANALYSIS OF COVARIANCE OF DELAYED TRANSFER TEST FOR 48 R&E AND 48 GD SUBJECTS USING THE INITIAL TRANSFER TEST DIFFERENCES, DAT SCORES, AND TIME ON THE PROGRAM AS COVARIANCES

Source of Variation	Sum of Squares	d.f.	Mean Squares	F
Between Treatments	89.03	1	89.03	0.888
Error (Within)	9128.22	91	100.31	
Total	9217.25	92		

F required for significance at .05 level = 5.20 (two-tailed test).

TABLE 7

ANALYSIS OF COVARIANCE OF RETENTION TEST FOR 48 R&E AND 48 GD SUBJECTS USING DAT SCORES AND TIME AS COVARIATES

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Between Treatments	23.95	1	23.95	.69
Error (Within)	3178.69	92		
Total	3202.65	93		

F required for significance at .05 = 5.29 (two-tailed test).

TABLE 8

ANALYSIS OF VARIANCE OF INITIAL TRANSFER TEST FOR 48 R&E AND 48 GD SUBJECTS DIVIDED INTO HIGH AND LOW ABILITY GROUPS BASED ON PRETEST SCORES

Source of Variation	Sum of Squares	d.f.	Mean Squares	F
Between Treatments	728.7	1	728.7	5.07
Between Abilities	4080.4	1	4080.4	28.38
Interaction	2.1	1	2.1	.01
Error (Within)	13230.1	92	143.8	
Total	18041.3	95		

F required for significance at .05 level = 4.00 (one-tailed test)

TABLE 9

ANALYSIS OF VARIANCE OF DELAYED TRANSFER TEST FOR 48 R&E AND 48 GD SUBJECTS DIVIDED INTO HIGH AND LOW ABILITY GROUPS BASED ON PRETEST SCORES

Source of Variation	Sum of Squares	d.f.	Mean Squares	F
Between Treatments	658.8	1	658.8	4.18
Between Abilities	1646.4	1	1646.4	10.44
Interaction	91.6	1	91.6	.58
Error (Within)	14509.1	92	157.7	
Total	16905.9	95		

F required for significance at .05 level = 4.00 (one-tailed test).

TABLE 10

ANALYSIS OF VARIANCE OF DELAYED RETENTION TEST FOR 48 R&E AND 48 GD SUBJECTS DIVIDED INTO HIGH AND LOW ABILITY GROUPS BASED ON PRETEST SCORES

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Between Treatments	0.4	1	0.4	.01
Between Abilities	227.1	1	227.1	5.70
Interaction	48.7	1	48.7	1.22
Error (Within)	3667.9	92	39.8	
Total	3944.1	95		

F required for significance at .05 level = 4.00 (one-tailed test).

TABLE 11

ANALYSIS OF COVARIANCE OF INITIAL TRANSFER TEST FOR 42 MALES AND 54 FEMALES USING DAT SCORES AND TIME AS COVARIANTES

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Between Treatments	330.07	1	330.07	2.59
Error (Within)	11722.55	92	127.42	
Total	12052.62	93		

F required for significance at .05 level = 5.20 (two-tailed test).

TABLE 12

ANALYSIS OF COVARIANCE OF DELAYED TRANSFER TEST FOR 42 MALES AND 54 FEMALES USING INITIAL TRANSFER TEST, DAT SCORES AND TIME AS COVARIATES

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Between Treatments	93.62	1	93.26	0.93
Error (Within)	9123.98	91	100.26	
Total	9217.25	92		

F required for significance at .05 level = 5.20 (two-tailed test).

TABLE 13

ANALYSIS OF COVARIANCE OF DELAYED RETENTION TEST FOR 42 MALE AND 54 FEMALE SUBJECTS USING INITIAL TRANSFER TEST DAT SCORES, AND TIME AS COVARIATES

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Between Treatments	25.65	1	25.65	0.74
Error (Within)	3176.98	91	34.91	
Total	3202.63	92		

F required for significance at .05 level = 5.20 (two-tailed test)

TABLE 14

ANALYSIS OF COVARIANCE OF INITIAL TRANSFER TEST FOR 28 SUBJECTS FROM SCHOOL 1 AND 68 SUBJECTS FROM SCHOOL 2 USING DAT SCORES AND TIME AS COVARIATES

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Between Schools	150.51	1	150.51	1.163
Error (Within)	11902.11	92	129.34	
Total	12052.62	93		

F required for significance at .05 level = 5.20 (two-tailed test).

TABLE 15

ANALYSIS OF COVARIANCE OF DELAYED TRANSFER TEST FOR 28 SUBJECTS FROM SCHOOL 1 AND 68 SUBJECTS FROM SCHOOL 2 USING INITIAL TRANSFER TEST DAT SCORES AND TIME AS COVARIATES

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Between Schools	62.02	1	62.02	0.616
Error (Within)	9155.23	91	100.61	
Total	9217.25	92		

F required for significance at .05 level = 5.20 (two-tailed test).

TABLE 16

ANALYSIS OF COVARIANCE OF DELAYED RETENTION TEST FOR 28 SUBJECTS FROM SCHOOL 1 AND 68 SUBJECTS FROM SCHOOL 2 USING INITIAL TRANSFER TEST DAT SCORES AND TIME AS COVARIATES

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Between Schools	2.25	1	2.25	0.064
Error (Within)	3200.39	91	35.17	
Total	3202.64	92		

F required for significance at the .05 level = 5.20 (two-tailed test).

TABLE 18

TESTS COMPARING MEAN SCORES OF SCHOOL 1 WITH MEAN SCORES OF SCHOOL 2 ON INITIAL TRANSFER, DELAYED TRANSFER, AND DELAYED RETENTION MEASURES

	School 1		School 2		t	p
	M	s	M	s		
Initial Transfer	50.75	15.81	43.80	12.42	2.04	.02
Delayed Transfer	44.20	15.06	42.10	12.63	0.64	.26
Delayed Retention	16.57	7.01	15.15	8.65	0.84	.20

(One-tailed test)

TABLE 20

COMPARISON OF CHANGE SCORES FROM THE INITIAL TRANSFER TEST TO THE DELAYED TRANSFER TEST FOR SCHOOL 1 AND SCHOOL 2

	School 1	School 2	t	p
Net Change	6.55	1.70	1.86	.03

(One-tailed test)

TABLE 19

ANALYSIS OF COVARIANCE OF ORAL PROBLEM SOLVING FOR 14 R&E SUBJECTS AND 14 GD SUBJECTS USING DAT SCORES AND TIME ON PROGRAM AS COVARIATES

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Between Treatments	3.42	1	3.42	0.151
Error (Within)	544.34	24	22.68	
Total	547.76	25		

TABLE 20

ANALYSIS OF COVARIANCE OF ORAL PROBLEM SOLVING FOR 14 R&E SUBJECTS AND 14 GD SUBJECTS USING THE INITIAL TRANSFER TEST, DAT SCORES AND TIME ON PROGRAM AS COVARIATES

Source of Variation	Sum of Squares	d.f.	Mean Square	F
Between Treatments	0.10	1	0.10	0.005
Error (Within)	458.75	23	19.95	
Total	458.86	24		

References

Cronbach, L. J., "Coefficient alphas and the internal structure of tests," Psychometrika, 16, 1951, 297-334.

Gagné, R. M. and Brown. L. T., "Some factors in the programming of conceptual learning, J. exper. Psychol., 62, 1961, 313-321.

Gagne, R. M. and Smith, E. C., "A study of the effects of verbalization on problem solving," J. exper. Psychol., 63, 1962, 12-18.

Guilford, J. P., Fundamental Statistics in Psychology and Education. New York: McGraw-Hill, 1956.

Hays, W. L., Statistics for Psychologists. New York: Holt, Rinehart and Winston, 1963.

Hendrix, G., "Learning by discovery," Math Teacher, 54, 1961, 290-299.

Kersh, B. Y., "The motivating effect of learning by directed discovery," J. educ. Psychol., 53, 1962, 65-71.

Kersh, B. Y., Directed Discovery vs. Programed Instruction. Final Report Title VII, Project Number 907. Oregon: Oregon System of Higher Education, 1964.

Richardson, R. C. and Pawar, S. B., Population Characteristics of the Salt Lake Metropolitan Area. Salt Lake City: University of Utah, Bureau of Economic and Business Research, 1964.

CHAPTER V

DISCOVERY VS. EXPOSITORY SEQUENCING IN SIX WEEKS OF CLASSROOM INSTRUCTION IN MATHEMATICS*

Blaine Richard Worthen

The study reported in the previous chapter was limited to a comparison of student achievement on self-instructional materials requiring only a few hours study. The present study covers a six week instruction period with teacher presentation of carefully sequenced materials, each teacher being required to teach one class using a "discovery" sequencing approach and another class using an "expository" sequencing approach.

INTRODUCTION

The purpose of this study was to provide research evidence, in a classroom learning situation, of the relative effectiveness of two methods of teaching selected mathematical concepts to fifth- and sixth-grade pupils. The methods included a discovery method and an expository method in which varying amounts and types of teacher guidance were employed and special instructional materials were used. The major difference between experimental Treatments D (discovery) and E (expository) was in the sequence characteristics of the presentations. In Treatment E, verbalization of the mathematical concept or generalization was the initial step in the instructional sequence by which the concept or generalization was to be taught; the subject received problems to solve that were instances of

---

This is an edited version of a thesis submitted by Blaine Richard Worthen to the Graduate School, University of Utah (June, 1965) in partial fulfillment of the requirements for the degree of Master of Science.

the principle only after the initial verbal and written presentation of the principle were given. In Treatment D, verbalization of the mathematical concept or generalization was delayed until the end of the instructional sequence by which the concept or generalization was to be taught; the subject received problems that were instances of the principle in verbal and written sequences structured to facilitate discovery of the generalization.

The criteria used to measure the outcomes of instruction included the following: tests of initial learning, retention, and transfer of the selected mathematical concepts; tests for transfer of heuristics; and measures of attitude toward the subject content.

Hypotheses

There is a between-treatment difference, in favor of Treatment D in: (1) initial learning, as measured by the mean scores on a criterion test of the selected mathematical concepts given immediately after instruction; (2) retention, as measured by the mean scores on a criterion test of the selected mathematical concepts given two weeks after instruction; (3) concept transfer, as measured by the mean scores on a transfer test of the selected mathematical concepts given immediately after instruction; (4) negative concept transfer, as measured by the mean scores on a negative concept transfer test given immediately after instruction; (5) attitude, as measured by the mean scores of attitudinal criterion tests given immediately after instruction; (6) the transfer of heuristics, as measured by the mean scores on two heuristic transfer tests given immediately after instruction.

Rationale for Hypotheses

Hypotheses 1, 2, and 3 are an outgrowth of the writer's personal interpretation of the equivocal results of the literature summarized in Chapter II. Although no concrete conclusions can be drawn from the hodgepodge of research which has been conducted on discovery learning, some plausible inferences can be made.

One such inference is that short-term studies are likely to favor expository instruction while long-term studies are likely to favor discovery learning. This inference is based on the assumption that pupils have typically been trained to learn from expository instruction; and hence need a relatively longer period of time in which to learn to utilize the techniques necessary in discovery learning. To the extent that this assumption is correct, the first three stated hypotheses are tenable in view of the length of this study.

A second inference that can be drawn is that many researchers have destroyed some of the power of their "discovery" treatment in an attempt to control nonexperimental variables. While control of many relevant variables has often been omitted, the control of many irrelevant variables has operated against the possibility of pupils experiencing real discovery in many of the so-called "discovery" treatments. Conversely, such procedures have left the expository treatment relatively uneffected. To the extent that this premise is true, this has acted to reduce the relative effectiveness of discovery learning in many of the previous studies. Admittedly, some studies have also discriminated against exposition by applying the term to methods of instruction which resulted in obvious rote learning. In the present study, an attempt is made to control the relevant variables but to do so in a way which allows each treatment to develop the optimum power of which it is capable. To the extent that this attempt is successful, hypotheses 1, 2, and 3 seem to be tenable.

Hypothesis four is an outgrowth of the writer's impressionistic observations of pupils in the initial tryout of the instructional materials. During this tryout, it seemed that pupils who were specifically warned against overgeneralizaing a principle (Treatment E) continued in overgeneralizing much longer than the pupils who found empirically (Treatment D) that the principle could not be applied in all situations.

Hypothesis five is also based on impressionistic observations in the pilot study, coupled with work by Kersh (1958, 1962) which suggested that any advantages of discovery learning could best be explained in terms of pupil motivation.

Hypothesis six is based on a logical extension of interpretations by Suchman (1962) and Della-Piana (1957) relating to searching set and its effect upon conceptual data processing.

Procedures

Sampling

The experimental sample was comprised of 432 pupils in 16 fifth- and sixth-grade classes in Salt Lake City School District, Salt Lake City, Utah, during the 1964-65 school year. An additional 106 pupils in three sixth-grade classes were used as a control group, bringing the total of 538 pupils used in this study. Total enrollment in grades five and six was 5,962 as of March 4, 1965.

Selection of Experimental and Control Groups. Random selection of the sample was limited by several factors. Chief among these was the necessity for Salt Lake City District central office personnel to decide which schools could be involved in the study. In making this decision, they were faced with some practical limitations. First, adherence to the research design eliminated from consideration all schools having less than three fifth- or sixth-grade classes. The research design also dictated the use of schools where two or more of these classes were taught mathematics by the same teacher. These requirements were necessary because of the design for each teacher to teach two experimental classes and also to have access to a third class to practice various teaching techniques. From the schools which met these organizational criteria, further selection was made on the basis of which principals were willing to use the experimental materials as the sole mathematics curriculum for these two classes for a seven-week period. Finally, it was necessary to determine which of the remaining schools had a teacher who met the suggested selection criteria. These suggested criteria were: (a) general competence in the classroom, (b) competence in the teaching of arithmetic, (c) three or more years of teaching experience, and (d) desire to participate in this research project. Central office personnel were asked to select a minimum of ten such teachers.

These selection procedures resulted in the selection of eleven teachers in eleven schools. Personal considerations prevented three of these teachers from participating in the study, although one of them did subsequently teach the three control groups. The remaining eight teachers (four male and four female) ranged in teaching experience from five to thirty years and taught in eight widely-scattered schools which were judged by central office personnel to be fairly representative of the various socio-economic and geographical segments of the population.

Four of the eight teachers taught arithmetic to three classes at the same grade level. In these cases, arithmetic computation, arithmetic problem-solving, and IQ scores were obtained for each pupil in each class.

Means for each class were computed and used as the basis for selecting as experimental classes for each of these teachers their two most comparable classes.

The subjects used in this study were those pupils who had been assigned to the two experimental classes of each of the eight teachers selected. Although this sample was not arrived at by random sampling, there was no basis for expecting these groups to differ on the relevant variables under consideration. Because of differences which did show up, however, it was necessary to introduce statistical controls in the analysis of data.

Of the 16 classes involved in the study, 14 were comprised of sixth-grade pupils. In one school, two fifth-grade classes were used. The inclusion of these fifth-grade groups seemed justifiable as there was no reason to believe that the efficacy of either method was dependent on the grade level of the learner. In addition, the original instructional materials had been written to be used with either the fifth or sixth grades.

Tables 1 and 2* include descriptive data on each of the 16 experimental classes used in this study plus the control classes.

Experimental Design and Controls

Two classes in each of eight schools served as experimental groups and three classes in a ninth school served as the control group. In each of the experimental schools, both classes were taught arithmetic by the same teacher, one class by Treatment D and one class by Treatment E. This was done in order to control the dimensions of teacher personality and other teacher characteristics. Seven of the teachers taught two sixth-grade classes while the eighth teacher taught two fifth-grade classes. In the control school, all three sixth-grade classes were taught arithmetic by the same teacher.

---

*All tables referred to in this chapter are presented together at the end of the chapter.

The experimental period was approximately seven weeks in length, consisting of three days of pre-test administration, a six-week instructional period, and five days of post-test administration. The control group received both the pre- and post-tests but received no instruction on the selected mathematical concepts during the intervening six-week period.

The six major non-experimental variables controlled in this study are presented below.

(1) The research design balanced, as nearly as possible, the number of classes taught within each treatment during any particular segment of the school day and the number of homeroom classes taught within each treatment.

(2) The pupils in Treatments D and E received the same length of time to work on the learning tasks. This was done by equating, between the two treatments, the length of the daily learning periods and the number of daily learning periods during the experimental instructional period. All instruction was terminated at the end of the six-week experimental period; in no case was an extension of time given to allow the teacher to complete the presentation of the unit. While teacher differences were expected to result from this procedure, it was found that there would be no significant treatment differences in amount of

material covered.

(3) Although the type of verbal behavior varied to fit the two teaching models, the amount of verbalization in the teacher's oral presentation and in the written instructional materials was held constant in both experimental treatments. Verbalization of the mathematical generalizations varied in sequence between the two treatments but was present in the instructional procedures used with both groups.

(4) In order to obviate the criticism that the instruction received by the two experimental treatment groups was not actually different or did not match the experimental model, three techniques were used in this study in an attempt to assess the extent to which the teachers did, in fact, teach by the specified techniques. These techniques included the following: (a) live rating of three lessons in each treatment for each teacher by observer-raters, (b) rating of three lessons in each treatment for each teacher from lessons recorded on audiotape, and (c) rating of teaching behavior on the discovery-expository dimension by pupils. A 20 per cent sample of the teaching behavior of each teacher in each treatment was rated by techniques (a) and (b) above, while the third technique was used as an indication of the teacher's total teaching behavior during the experimental period.

(5) The research design and all of the various procedures and methods utilized were calculated to negate any differential "Hawthorne Effect" between the two experimental groups. It is possible, however, that the control group was at a disadvantage as far as teacher and pupil enthusiasm were concerned since the control teacher did not participate in the training sessions nor use any "new" materials. The control group in this study was simply a base-line "no-treatment" group.

(6) An attempt was made to equalize the pre-experimental mathematical experiences of all subjects in Treatments D and E in order to allow both groups a common background from which to draw in the acquisition of the selected mathematical concepts. A "pre-experimental" unit including both specific and general mathematical concepts was judged to provide necessary background for the experimental materials. This unit was conducted in each class during a two-month period immediately preceding the pre-tests. In addition, pollution of the experimental results by non-experimental arithmetic experiences was minimized by a requirement that no homework or out-of-school arithmetic assignments be given to the pupils. District personnel complied with this request and also elicited parental cooperation.

Training Program

A twenty-week training program was conducted which was attended regularly by all raters and experimental teachers. This training class met a minimum of two hours weekly for twenty weeks. Extra sessions of longer duration were inserted into the training program as they proved necessary. This program ran for 13 weeks prior to the experimental period and seven weeks concurrently with this period. Training was given in three areas. First, all teachers and raters were trained in a body of general mathematical concepts which were developed to act as background for the specific concepts used in the research project. In addition, thorough training was given in the use of all selected mathematical concepts used either in the instructional period or in any of the criterion measures. Second, the teachers were thoroughly trained in the procedures of administering and scoring the various tests, scales, and questionnaires. Third, the teachers and raters were thoroughly trained in the use of the two specific methods of instruction under consideration in this study. A detailed description of these two methods is provided in the section "Instructional Procedures, Teaching Models and Techniques." Training procedures in the use of the two methods of instruction included demonstrations by the experimenter of all instructional units in each treatment, practice teaching and critiques of small portions of the instructional units during the training class, and practice of instructional techniques in a third class set up in each school specifically for practice purposes.

Instructional Materials

The instructional materials used by the subjects in both treatment groups were written by the experimenter and included the presentation of several mathematical concepts specifically selected for this study. These concepts were selected on the basis of two considerations: (a) their suitability for both discovery and expository teaching; and (b) uniqueness and resultant lack of pupil proficiency on the concept at the beginning of the investigation.

The instructional materials were first tried out in pilot study conducted with two sixth-grade groups at the William M. Stewart School during the 1963-64 school year. A revision of the initial materials was again tried out in one fifth- and sixth-grade class at the William M. Stewart School in the 1964-65 school year. The results of the 1964-65 tryout shaped the final revision of the materials into the form in which they were used in this study and in which they appear in Appendices I and II published as Part II of this final report in a separate volume.*

---

*All Appendices referred to in this volume are presented in Part II of the Final Report published as a separate volume.

These materials were equated in terms of the mathematical concepts, diagrams of physical models, number and type of examples, and degree of verbal presentation used for each treatment. The two sets of materials differed primarily in terms of sequence characteristics.

Credit should be given to sources from which the experimenter gleaned ideas which were utilized in the construction of the materials. Ideas from at least the following three sources have been adapted into the instructional materials written for this experiment: University of Illinois Committee on School Mathematics, High School Mathematics: Unit 1, The Arithmetic of Real Numbers, University of Illinois Press, Urbana, 1962; Wirtz, R. W., Botel, M., and Sawyer, W. W., Math Workshop for Children, Levels D and E, Encyclopaedia Britannica Press, Inc., 1962; School Math Study Group, Mathematics for the Elementary School, Grade 5, (Part 1), revised edition, 1962.

The materials were written in such a way that they could be given to a pupil who had missed a learning session with some confidence that he could work through that segment of the materials independently and gain at least the rudiments of the concept. Typically, however, the materials were not used in that way.

The materials for each treatment were subidvided into six concept sections. Each of these sections was labeled by a code abbreviation. While this code merely served to differentiate the sections in Treatment E, it was necessary in Treatment D in order to avoid cues that would aid verbalization prior to the pupils' discovery of the concept. Lists of these abbreviations and the corresponding topics for each concept section appear at the beginning of both Appendix I, page A1.01 and Appendix II, page A2.01.

Five pages were deleted from the materials written for Treatment D (Appendix I, pages A1.40 and A1.44, inclusive) when the experimenter found he was unable to produce a comparable section for the expository treatment which was meaningful and yet did not allow self-discovery. While these pages were not used in the study, they had previously been printed and are included in Appendix I in order to complete the pagination.

Instructional Procedures

The sequence of presentation was determined by the instructional materials used in both treatments. These are shown in Appendices I and II. Both treatments consisted of initial verbal introduction of the unit by the teacher. In Treatment D, this usually consisted of teacher explanation of a pseudo-physical model (pump-tank-film, etc.) or presenting pupils with a sequence of computational problems through which a discovery could be made. In Treatment E, the initial teacher introduction generally consisted of verbal and symbolic explanation of the mathematical concept taught. The printed pages which dealt with the same concept were not introduced in either treatment until after the initial blackboard and oral presentation by the teacher. Pupils were told that they could use these pages to review the model or concept which had been presented. The pupils were then given the problem pages and were left to refer to the explanatory pages as they saw fit. These pages were given to the pupils to serve for review purposes and were also utilized as the sole instructional content received by those who had been absent during that particular learning period. In subsequent make-up periods, the pupil used these explanatory pages as a primary source of instruction, with the teacher giving help only as needed.

Rate of Presentation. The rate of instruction and progress through the instructional materials varied among the teachers. This variation was a result of the varying IQ and math achievement levels among the classes as well as the normal tempo of presentation of each teacher. In order to insure adequate presentation of the concepts to both treatments, a criterion was established that a minimum of 85 per cent of the treatment group must attain a minimum level of understanding of each concept before the teacher was allowed to proceed to the next concept.

The teachers were not required to complete all six of the instructional units contained in each set of materials. The only expectation was that they would go through the concepts and the materials carefully, making every attempt to make both treatments meaningful. When the final day of the experimental period came, each treatment group was terminated wherever the deadline found it. The pupils were told that they would have an opportunity to conclude the unit later in the year. This was delayed, however, until after the retention test was given. Table 3 shows the page on which each treatment group terminated.

Individual Differences and Help. Most of the classes were reasonably heterogeneous in IQ and arithmetic ability, and consequently pupils varied widely in the speed with which they completed the assignments. Because of this factor, free-reading materials were provided in each class for the use of pupils who finished the assigned work quickly.

In order to allow the teachers additional time to help those pupils who were having difficulty attaining some minimal understanding of the concept, occasional days were specified for individual help in both treatment groups. On these days, the majority of the pupils in the class occupied their time with free-reading while the teacher gave individual assistance to those who needed it. In addition, each teacher was allowed to give individual help above and beyond normal class time wherever this was feasible, as long as this opportunity was given equally to pupils in both treatment groups.

Teaching Models and Techniques

Five dimensions of teaching behavior were used as the basis for constructing instructional models for Treatments D and E. These were selected from among many characteristics which appeared to differentiate discovery from expository teaching but which could not be readily identified in a teacher's classroom behavior. All five of the selected dimensions were observable in some degree in most classroom teaching. These five characteristics also constituted the basic dimensions of the teacher rating scale. Description of the manner in which these characteristics were rated on the rating scale is given in "Use and Scoring--olR, taR" on page A3.67 of Appendix III.

The following section specifies model teaching behavior for each treatment on each of the five dimensions.

Treatment D. Model "discovery" teaching behavior on each of the five characteristics is briefly outlined here along with specific teaching techniques which are related to each item.

(1) Teacher onmiscience: In the discovery method, the teacher makes an effort to avoid any behavior which projects an onmiscient teacher attitude to the pupils and that consequently might deter from them questioning the correctness of the teacher's responses. The teacher acts as a resource person upon whom the pupils can depend for some measure of mathematical information. This information is given to the pupils in a "Let's see, I'm not quite sure . . . Why don't we check this and find out?" manner rather than by direct exposition.

The teacher who truly follows this treatment model does not give immediate evaluation of pupil responses but rather allows sufficient time for each pupil or other class member to discover and correct any error. If, after a time, the error has not been discovered by the class, the teacher appears to notice it for the first time and draws attention to it in a casual manner. When answers are given, they are checked for computational accuracy and the teacher studiously avoids any action which hints that he knows a "shortcut" or generalization which would enable more simple solutions to the problems. This "purposeful ignorance" can be used as a method of bringing the pupils to an awareness that a shortcut might be possible. This is done by use of the following technique. The teacher waits for behavioral evidence that a pupil has discovered the principle. This evidence is given by the pupil working the problems with such speed that the teacher can be virtually certain that discovery of the "shortcut" has taken place. The teacher then acts amazed at the pupil's speed and makes comments such as "My goodness, you certainly do multiply rapidly. I didn't know it was possible to multiply that fast." The teacher also makes certain that the pupil does not tell the class that he is using a method other than that specified in the problem, although the teacher's remark does imply to the class that other approaches might be possible.

(2) Introduction of generalization: In Treatment D, the teacher delays the verbalization of the concept until a minimum of 85 per cent of the pupils have made the discovery. Prior to this time, there is no hint that a principle exists, other than the indirect implication mentioned in the preceding section. At the end of the unit, after at least the minimum required number of pupils have made the discovery, the principle

is verbalized. This verbalization is not stated by the teacher. Rather, the teacher draws statements of the "shortcuts" from the class and then helps to refine these statements into reasonably precise formulations of mathematical ideas. Following this verbalization, no further practice on the principle is given.

(3) Method of answering questions: When a question is asked in a Treatment D class, the teacher answers it by referring to the model or taking the pupil back to the computational sequence which is to lead to self-discovery of the principle. When a pupil says, in essence, "I don't get it," or asks some other nontrivial or noncomputational question, the teacher avoids giving direct information on how to solve the problem. Instead, the teacher goes back to the model or to the problem sequence and asks the pupil questions about the model. The teacher returns to the model again and again, but never gives any hint that there is an underlying principle which can be discovered. The teacher is allowed to regroup the examples in a way which tends to draw the pupil's attention to the principle, but no verbal hint of the principle is given. Each pupil is allowed to discover the principle at some level. If, in the final analysis, the pupil can still not make the discovery, he is not told until such time as the verbalization of the concept is drawn from the class.

(4) Control of pupil interaction: In Treatment D, techniques are employed to insure that each pupil works independently and does not share his discovery with other pupils. Partial structuring for "not telling" is accomplished in the general administrative statement read to subjects in both treatments. An extra section of this statement, which emphasizes the necessity of not sharing information with other members

of the class, is read only to the discovery pupils. An occasional reminder of this by the teacher is generally sufficient to discourage such sharing. If a pupil does reveal the principle prematurely, the teacher expresses disapproval for the violation of the previous agreement. Provisions are made, however, so that pupils can share their discovery of the principle with the teacher in an acceptable manner. These techniques include: (a) having the pupils write the answer while the teacher circulates and looks at the written response, and (b) occasionally having the pupils respond silently by "mouthing" the answer, pointing, etc.

(5) Method of eliminating false concepts: In Treatment D, the teacher gives the pupil no information concerning the possibility of overgeneralizing a concept to problems on which it is not applicable. Rather, the teacher includes "trap questions" which resemble problems which can be readily solved by a particular principle. If and when the pupil discovers that his "rule" will not apply in this instance, the teacher gives no hint of why the rule did not work; rather, he allows the pupil to discover the necessary rationale.

Treatment E. Model "expository" teaching behavior on each of the five characteristics is briefly outlined here, along with specific teaching techniques which are related to each item.

(1) Teacher omniscience: In the expository method, the teacher makes an effort to project to the pupils an attitude of reasonable confidence in the teacher's knowledge of mathematics. He acts as the primary source of mathematical information and indicates that he is always able to demonstrate solutions to the majority of problems with which the pupils will work. On many occasions, he allows the pupils to solve the problems and turns questions back to them for answers. This is not done, however, with any degree of uncertainty on the teacher's part but is used

to give the pupils an opportunity to more fully understand the problem material. The teacher who follows this treatment model gives an immediate (although not negative) evaluation of the correctness of all responses. The teacher always checks the pupils' answers by use of the principle and continually illustrates the relation between specific examples and the principle by which they may be solved.

(2) Introduction of the generalization: The "expository" teacher sees that the generalization is given before the pupils receive any exposure to the instructional materials. These rules, stated first by the teacher, are then illustrated by examples. Relationships between the principle and the examples are constantly stressed.

(3) Method of answering questions: When a question is asked in a Treatment E class, the teacher answers it by relating it back to the rule, then giving further examples which clarify the way in which the rule is used to solve that particular type of problem. If a pupil says, in essence, "I don't get it," or asks other nontrivial or noncomputational questions, the teacher again verbally and symbolically explains and demonstrates the rule by the use of appropriate examples. The teacher is free to restate the rule in less technical language if, in his opinion, this may be helpful to the pupil. Care is taken, however, to maintain a reasonable degree of mathematical precision and to make no statement which would encourage self-discovery of the rule.

(4) Control of pupil interaction: In Treatment E, no attempt is made to control pupil interaction above and beyond that which would normally be expected in the classroom. Pupils are asked to comply with normal expectations of doing their own work and interacting only in constructive ways with other pupils. Pupils are free to ask any questions

with which they need help or state any principle which they feel will be helpful. Even though few limitations are placed on verbal interaction in this treatment, the teacher may still want to use techniques to test the understanding of the individual pupils quickly. These techniques could include: (a) having the pupils write the answer while the teacher circulates to examine it, and (b) having the pupils respond silently by "mouthing" the answer, pointing, etc.

(5) Method of eliminating false concepts: In Treatment E, the teacher warns the pupils of all of the common errors that are made in applying the principle. Specific types of "trap questions" are shown to the pupils and the inappropriateness of applying the principle to each is illustrated. Pupils are then given samples of such "trap questions" interspersed with problems in which the principle is applicable and are told to watch for the "trap questions" in solving the problems.

Sample Lessons in Treatments D and E. In Appendix IV, sample typescripts of audio-tape transcriptions are given of four lessons. The first two lessons were presented by the same teacher, one in Treatment D and one in Treatment E. These lessons were judged by the raters to be representative of lessons taught by one teacher in two different treatment models at the same point in the lesson sequence. Lesson 1, Treatment E, Teacher 1 is presented on pages A4.05 to A4.07. Lesson 1, Treatment D, Teacher 1 is presented on pages A4.08 to A4.11.

Excerpts from two lessons judged by the raters to be representative of the teaching in the two treatments is also included in Appendix IV.* Lesson 2, Treatment E, Teacher 8 was judged to be typical of the expository teaching which took place in Treatment E. Excerpts from this

---

The reader is reminded that all Appendices appear in a separate volume which is Part II of this final report.

lesson appear on pages A4.12 to A4.14. Lesson 2, Treatment D, Teacher 7 was judged to be typical of the discovery teaching which took place in Treatment D. Exceprts from this lesson appear on pages A4.15 to A4.17.

Instruments

Samples of all the instruments used in this study are included in Appendix III, with the exception of the three standardized IQ and mathematics achievement tests. The method of administration and scoring is included with each instrument. A list of the abbreviations used for each instrument appears on page A3.02. The abbreviations for each instrument are also included in the brief description of each instrument given in this section. Reliabilities and intercorrelations for the instruments are presented in Tables 4, 5, 6, and 7.

Attitude Measures

Three attitude measures were employed in this study. One was a test of actual behavior toward arithmetic; the other two were attitude scales.

cbAT - Choice Behavior Attitude Test. This instrument was intended to measure the subjects' attitude toward arithmetic by determining the degree to which the subjects selected an arithmetic task, as opposed to other tasks, in an actual choice situation.

This instrument was an adaptation of an instrument used by Umemoto and Haslerud (1955) in a comparative study of attitudes of American and Japanese children toward number situations.

Although the preliminary test-re-test (consecutive days) reliability check of this instrument yielded a reliability coefficient of .91, this reliability dropped to .13 in the subsequent (six week) test-re-test check. This was thought to be partially due to a phenomenon which occurred in the control group and in several experimental classes. It was observed that, in the re-test session, many pupils made the choice on the basis of "doing the side that I didn't do before." This reaction was found to occur to a marked extent in several classes. Test-re-test gains checked across all three treatments indicated that this "mirror-like" change from pre- to post-test choices was equally operative in both directions. Because of this phenomenon it was thought that this test may not be appropriate in a test-re-test situation. (In Umemoto and Haslerud's study, this test was administered to each subject only once.)

In addition to the above difficulty, an error in printing resulted in errors in administering this test in three schools. This, combined with the earlier dilemma, served to negate the use of this instrument in this study. Therefore, the data from this test were not analyzed. The instrument looks promising, however, and will likely be used by the writer in a subsequent study.

A sample of this instrument and an explanation of its administration and scoring are contained in pages A3.14 to A3.22 of Appendix III.

sdAT - Semantic Differential Attitude Test (scale). This instrument was intended to assess that aspect of pupil attitude which is reflected in reactions to polar adjuectives given in relation to arithmetic.

This scale was constructed along the principles of semantic differentiation developed by Osgood, Suci, and Tannenbaum (1958). Nineteen scales were included of which only 13 were used in scoring. The six scales which were not used in scoring were inserted to obscure the intent of the instrument. These six items were reported as having generally low loadings on the evaluative factor. All 13 of the scales used in the scoring were scales which received repeated high loadings on the evaluative factor in several of Osgood's studies (Osgood, et al, 1958, Chapter 2).

The adaptation of Osgood's scale used in this study was a five-point scale which was scored from negative two to positive two with a pivotal scoring of zero. The polar adjectives were arranged in such a fashion that the positive and negative scoring of the scale rotated. This was done to avoid any possibility of the pupils developing positional preferences.

A sample of this instrument and an explanation of its administration and scoring appear in pages A3.23 to A3.27 in Appendix III.

asAT - Attitude Statement Attitude Test (scale). This instrument was intended to measure that aspect of pupil attitude reflected in expressions of agreement or disagreement with both favorable and unfavorable statements concerning arithmetic.

This instrument was adapted from a similar instrument used by Umemoto and Haslerud (1955) and consisted of a five-point scale which was scored from negative two to positive two with a pivotal rating of zero. The oral statements were sequenced in such a fashion that the scoring on the ends of the scale rotated to avoid any possibility of the scoring being influenced by positional preferences of the pupils.

A sample of this instrument and an explanation of its method of administration and scoring are contained in pages A3.28 to A3.24 in Appendix III.

TOT-AT - Total Attitude Test. Each pupil's scores on asAT and sdAT were summed to yield a total attitude score. Since the two scales were judged to measure slightly different dimensions of attitude (Pearson r between tests was .54 on the pre-test and .68 on the six week post-test) this total score was thought to be the best measure of total attitude. This score was used in the analysis of the data in addition to the sdAT and asAT scores.

Measures of Teaching Behavior

Two instruments were designed in an attempt to assess the degree to which the teachers followed the expository or discovery models in their teaching. One was a questionnaire administered to all pupils; the second, a scale used by raters to rate teaching behavior.

stR - Student Rating of Teacher. This instrument was constructed in questionnaire form and included five distinguishing characteristics of teaching behavior which were used to identify expository and discovery teaching. These characteristics are discussed in the section on "Instructional Procedures, Teaching Models and Techniques." The questions were phrased in such a way that the pupils were likely unaware of the intent of the instrument.

This scale contained 25 items which could be subdivided on the basis of two classifications. First, each item could be classified on the basis of which of the five "teaching-behavior characteristics" it represented. In addition, the items could be classified into items which, if answered affirmatively, would typify "low discovery" teaching behavior, and items which, if answered affirmatively, would typify "high discovery" teaching behavior.

A sample of this instrument and an explanation of its method of administration and scoring are shown on pages A3.35 to A3.39 in Appendix III. The scoring system was adapted from a similar pupil questionnaire reported by Shaw (1963, p. 3).

olR, taR - Observer Live Rating, Tape Rating. A rating scale was developed which was used by raters in rating teachers in the actual classroom (olR) as well as by a rater who rated teachers from audio-tape recordings of the classroom lessons (taR).

Three persons served as raters. None of these persons taught in this experimental study. One rater, an experienced female classroom teacher, rated almost all of the auido-tape transcriptions. The other two raters served as raters in the actual classroom setting. One of these raters was the experimenter. The second was a female undergraduate honor student who was recommended by professors in education and mathematics. All raters were thoroughly trained in the selected mathematical concepts and in both treatment models.

The per cent of rater agreement on this instrument was defined as the per cent of the total pairs of ratings assigned by two or more raters rating the same teacher on the same lesson which were in perfect agreement. Multiple rating of a lesson was accomplished in three ways: (a) instances when both observer-raters simultaneously visited the same classroom; (b) instances when one of the observer-raters rated a lesson which was also audio-taped, transcribed, and rated by the tape-rater; and (c) instances when a tape transcription was rated by all three raters. The per cent of rater agreement obtained from these methods is reported in Table 7.

This rating scale and an explanation of its use and scoring are recorded in pages A3.66 to A3.73 in Appendix III.

TOT-CT - Total Concept Test

The intent of this instrument was to measure pupil achievement on the mathematical concepts contained in the instructional materials. This test was constructed by the experimenter and contained four subsections delaing with the four main concepts covered in the instructional materials. The subsections were as follows:

dpCT - Distributive principle concept test

eCT - Exponent concept test

rmCT - Real number multiplication concept test

raCT - Real number addition concept test

All four subsections of this test and an explanation of their administration and scoring are shown on pages A3.03 to A3.13 in Appendix III.

TOT-RT - Total Retention Test

This test was a parallel form of the concept test mentioned in the preceding section and was subidivided into four similar subsections with corresponding abbreviations.

This test was not given in the pre-test series in the control group. It was given in the post-test series, however, and these scores were correlated with the scores on both the pre- and post-test administrations of the TOT-CT. These are given in Table 4, with the TOT-RT, TOT-CT pre-test coefficient reported under "six-week" and the TOT-RT, TOT-CT post-test reported under "consecutive days."

The method of administration, scoring, and samples of the four retention subtests are shown on pages A3.40 to A3.50 in Appendix III.

Transfer Measures

Four instruments were used to measure transfer of various concepts and behaviors. A brief description of each is included in this section.

cTT - Concept Transfer Test. This instrument was constructed in an attempt to assess the degree to which the two treatments may influence the pupils' ability to transfer the concepts learned during the instructional period to dissimilar situations and to the solution of dissimilar problems.

This instrument and an explanation of its administration and scoring are shown on pages A3.51 to A3.56 in Appendix III.

cnTT - Concept Negative Transfer Test. Five items were included in the cTT which were then pulled out and scored separately. These five items were intended to assess any differences between the two treatments in terms of rigidity and fixation and in applying the principle, even to the point of applying it where it is inappropriate.

The five items which were separated from the body of the cTT are indicated on page A3.56 in Appendix III. Scoring for these items is also given on that page.

nsTT - Number Series Transfer Test. This test was constructed by the experimenter in an attempt to assess the effects of the two treatments upon transfer of heuristics on a "paper and pencil" task. The test was designed to test the degree to which the two treatments effect the ability of the pupils to use discovery techniques in solving new problems or discovering a new principle.

It was impossible to obtain a reliability coefficient by test-retest procedures with the same subjects because discovery of the principle in the first administration would invalidate the second administration of the test. Because of the nature of the instrument, it was impossible to obtain a split-half reliability estimate or any of the other usual reliability measures. The test was, however, administered to a total of 54 fifth- and sixth-graders at the William M. Stewart Laboratory School, University of Utah. Pupils in the two groups were matched as nearly as possible on IQ and arithmetic achievement scores, resulting in 27 pairs of pupils. The nsTT scores of the matched pupils in the two classes were then correlated, with a resulting correlation of .69. This was used as a crude estimate of reliability of this instrument.

This instrument and an explanation of its method of administration and scoring appear on pages A3.57 to A3.61 of Appendix III.

apTT - Associative Principle Transfer Test. This instrument was constructed by the experimenter in an attempt to measure the effect of the two treatments on transfer of heuristics on a task presented by the teacher through oral presentation. The entire problem sequence was conducted through oral presentation, although the six criterion problems were answered with paper and pencil.

This instrument was subject to the same limitations and difficulties concerning reliability checks as was the nsTT. The same procedure was used in an attempt to estimate the reliability of this instrument and resulted in a crude reliability estimate of .61.

This instrument and an explanation of its method of administration and scoring are contained on pages A3.62 to A3.65 in Appendix III.

Measures of Group Comparability

Scores on two standardized tests were collected for each pupil used in this study. These were, with one exception noted below, tests administered by the Salt Lake City School District personnel as part of their normal sixth-grade testing program. The scores obtained on these instruments were used to check the comparability of the treatment groups and as covariates in the analysis of covariance.

IQ Measure. The IQ test used in this study is the Pintner Intermediate Test: Form A (Harcourt, Brace, and World, 1938). The mean for this test in Salt Lake City was 106.4, with a standard deviation of 15.7. This test was given at the beginning of the current school year to all Salt Lake City pupils in the sixth grade.

One exception to the above is noted in the school where two fifth-grade groups were used. Here, the IQ scores were from the Pintner-Durost

Elementary Test: Scale 2 administered at the third-grade level (Haroucrt, Brace, and World, Inc., 1940).

Arithmetic Achievement Measures. The two arithmetic achievement scores used in this study were obtained from "Test 5, Arithmetic Computation" and "Test 6, Arithmetic Problem Solving and Concepts," which are two sections of the Metropolitan Achievement Tests (Harcourt, Brace, and World, Inc., 1958). Salt Lake City District means on these two scores are 6.3 for "computation and 6.8 for "problem solving and concepts."

An exception to the above is again noted in the two fifth-grade classes used in the study. Here, the scores were obtained from the "Arithmetic Computation" and "Arithmetic Reasoning" sections of the Stanford Achievement Test: Form K, Intermediate Battery (World Book, Inc., 1953).

Statistical Treatment of Data

Data for this study were generated from three primary sources: (a) ratings of teaching behavior; (b) measures of attitude, initial learning, transfer, and retention on instruments created for this study; and (c) measures of IQ and math achievement on standardized tests.

In order to form groups with equal N's to facilitate statistical analysis of these data, the scores of 35 pupils (randomly selected within 11 classes needing reduction) were excluded from analysis. In four classes a total of 16 cases were added at the mean. These adjustments resulted in 16 classes, each with an equal N of 27, for purposes of data analysis. The adjustments made are shown in Table 8.

Means and Standard Deviations. On each instrument, the raw scores of each pupil were summed and means and standard deviations for each class were computed. The four subtests (dpCT, eCT, raCT, and rmCT) of the concept test (TOT-CT) were not considered as separate instruments. Pupil scores on these subtests were summed and the total (TOT-CT) score for each pupil was used in the data analysis. The same procedure was followed with the retention test (TOT-RT) which was a parallel form of TOT-CT. The pupil scores were then summed again, across all eight classes in each treatment, and the treatment means and standard deviations were computed for each instrument.

In addition to computing means and standard deviations for each class and treatment on each of the two attitude measures (sdAT and as AT), pupil scores on each of these instruments were summed to yeild total attitude scores (TOT-AT) for which means and standard deviations were also computed for each class and treatment. These data were used in the covariance analysis as well as data from the sdAT and the asAT.

Analysis of Covariance. The statistical technique used for making a majority of the comparisons was the standard analysis of covariance. Two separate analyses of covariance were used to analyze the data contributed by the majority of the criterion measures: (a) a two-way teacher-by-freatment analysis of the data generated within the two experimental groups, and (b) a simple analysis using treatment (including Treatment C) as the main effect.

The choice of covariates was determined by an examination of the intercorrelations on all measures and variables. On this basis, IQ, arithmetic computation, and arithmetic problem solving were used as constant covariates in the analysis of each dependent variable. (Sex differential, which in many studies has been found to be highly related to performance, was found to be unrelated to any of the measures used in this study, and therefore, was not used as a covariate.) In addition, pre-test scores were used as additional covariates in analysis of the post-test of each instrument administered in both the pre- and post-test series. These tests included asAT, sdAT, TOT-AT, and TOT-CT. TOT-CT post-test scores were used as an additional covariate in the analysis of the TOT-RT and cTT.

Analysis of Variance. The data generated from two instruments, stR and olR-taR, were analyzed independently from the data contributed by other measures.

The olR-taR data were analyzed in the following manner. First, a mean of all ratings assigned to each teacher in each treatment on every rating item was computed. This resulted in each teacher having a mean rating on each item on the rating scale in both their Treatment D and Treatment E ratings. These mean values were then separated into two general categories. The first category contained items one through five, those items thought to differentiate between expository and discovery teaching. The second category contained items six through eleven, the "general teaching" items. Three analyses of variance were used within each of these vategories: (a) an analysis which compared Treatment D ratings among teachers; (b) an analysis which compared Treatment E ratings among teachers; and (c) an analysis which compared total Treatment D with Treatment E ratings. A fourth analysis was used for the items one to five comparing the Treatment D and E difference scores between actual ratings and the theoretically perfect rating for each treatment.

The stR data were analyzed in the following manner. First, the index was computed for each pre- and post-test item for each teacher in each treatment. (For a description of this technique see the section on "Scoring - stR" on page A3.39 of Appendix III.) A pre- to post-test gain for each item for each teacher in each treatment was computed, and these values were used in three separate analyses of variance. The first of these compared Treatment E gains among teachers, and the second compared Treatment D gains among teachers. The third analysis compared total Treatment E with total Treatment D gains.

Whenever comparisons among teachers on either the olR, taR or stR resulted in a significant F ratio, the Scheffé S-method (Scheffé, 1959) was used to test which of the between-teacher comparisons were significant.

t Ratios. The t statistic was used for two purposes: (a) to assess the comparability of the three treatment groups (D, E, and C) on the IQ, math achievement, and pre-test measures; and (b) to test differences between the adjusted treatment means in instances when the simple between-treatment analysis of covariance yielded a significant F ratio.

RESULTS AND DISCUSSION

In the preceding sections, the problem under consideration has been stated, hypotheses pertinent to the problem have been formulated, and specific procedures for testing these hypotheses have been outlined. The results of this investigation and a discussion of these results are presented here. All tables relating to these results are presented in a section at the end of the discussion.

The Control Group

The investigation of the two methods of presentation (Treatments D and E) was based on the hypothesis that the independent variables included in the presentation of the selected mathematical concepts by the different methods would result in different learning outcomes in the two treatment groups. Another hypothesis, although not explicitly stated, was that both the discovery and expository methods would result in greater learning than no instruction. Therefore, the control group was included to test this hypothesis as well as to estimate the amount of learning attributable to receiving both the pre- and post-test series of the criterion measures. The control group was also used to assess the effects of other experimentally uncontrolled learning such as prior knowledge of the selected mathematical concepts, interaction among pupils between test periods, and independent reflection and study. Although efforts were made to control these variables, it was recognized that the possibility of some experimentally uncontrolled learning could not be completely precluded.

The performance of the pupils in each experimental group was compared with the performance of the pupils in the control group on all but two of the criterion measures. The two exceptions resulted from an occasion when a substitute teacher taught the control classes. The substitute teacher inadvertently exposed the pupils in the control group to two mathematical concepts employed in the nsTT and apTT, thus invalidating the use of these instruments with this group.

Comparability of Treatment Groups

In order to facilitate intertreatment comparisons of the initial pre- and post-test means (prior to adjustment by the covariance technique), Table 9, which shows the initial total treatment means and standard deviations on all instruments, is included for reference.

A pre-treatment inspection of the mean values for each treatment group was conducted on IQ, arithmetic computation, arithmetic problem-solving, and all pre-test measures. Although this inspection of pre-treatment group comparability revealed significant differences between pupils in the control group and pupils in one or both of the experimental groups on several of these measures, the only significant differences found between the Treatment D and the Treatment E groups were on the attitude measures. Pupils in Treatment E showed, in the pre-treatment comparisons, a significantly better attitude toward arithmetic than pupils in Treatment D. Table 10 shows the t values obtained in inter-treatment comparisons on each of the measures used to assess pre-treatment group comparability.

The ability and pre-test variables, with the exception of the stR pre-test, had positive and significant correlations (Pearson r) with the scores on the criterion measures. Because of the noncomparability of the three treatment groups on the IQ, arithmetic achievement, attitude pre-test, and concept pre-test measures, as well as for other considerations, the analysis of covariance technique was employed in analyzing the data contributed by these measures. The simple analysis of variance was used to analyze the data contributed by the stR.

Analysis of Teaching Behavior

Many investigations in which each teacher presented the learning task by two or more methods have been subject to the criticism that the teachers were unable to vary their teaching behavior sufficiently to effect a real test of the various treatment models. Procedures followed in this study to assess teacher fidelity to the two teaching models are presented below.

Analysis of Teaching Behavior as Measured by olR, taR

A 20 per cent sample of the teaching behavior of each teacher in each treatment was rated by two techniques: (a) rating by observer-raters of three classroom presentations, and (b) rating by a tape-rater of three audio-tape transcriptions of classroom presentations.

These ratings on the olR, taR scale were subdivided into two categories: (a) ratings on those items used to differentiate between discovery and expository teaching (items 1 to 5), and (b) ratings on those items used to assess general teaching behavior (items 6 to 11).

Items 1-5, Discovery - Expository Items. Comparison of the mean ratings for each teacher in each treatment on items 1 to 5 were used to determine teacher fidelity to the two prescribed teaching models. Perfect adherence to the discovery model would have theoretically resulted in a mean teacher rating of 5.0, while perfect adherence to the expository model would theoretically have resulted in a mean teacher rating of 1.0. Table 11 shows the mean rating by teacher and treatment for the first five items in the scale.

Two analyses of variance were carried out on this data. The first analysis of variance compared mean teacher ratings in Treatment D with mean teacher ratings in Treatment E and yielded a highly significant difference between the two experimental treatments. This analysis confirmed the notion that the instruction given to both groups was consistently dissimilar. Table 12 presents the results of this analysis.

The second analysis of variance compared the proximity of the obtained mean teacher ratings in each treatment to the perfect score for each of the theoretical teaching models. No significant differences were found in this comparison, thus substantiating the idea that instruction in both treatments followed the prescribed teaching models equally well. The results of this analysis are shown in Table 13. The relatively high overall per cent of fidelity to the teaching models was judged by the experimenter to be a function of both the construction of the instructional materials to fit both models of teaching and the rather specific teaching acts and behaviors rated by the olR, taR scale.

The mean teacher ratings on items one through five were also used to compare teacher differences within treatments. One analysis of variance compared the mean ratings for each teacher in Treatment D, while the second analysis compared the differences among the mean teacher ratings in Treatment E. No significant differences between teachers were found with either of these analyses. Tables 14 and 15 present the results of these comparisons.

Items 6-11, General Teaching Items. Comparisons of between-treatment differences in the mean teacher ratings on items 6 to 11 were used to determine the degree to which the teachers differed in their "general teaching" in the two treatments. The highest possible rating on each of these items was 5.0. The means ratings, by teacher and treatment, for items 6 to 11 are shown in Table 16.

One analysis of variance was used to compare mean teacher ratings in Treatment D with mean teacher ratings in Treatment E. This analysis showed no significant differences between the mean teacher ratings in the two treatment groups. Table 17 presents the results of this comparison.

Teacher differences within treatments on items 6-11 were also analyzed. One analysis of variance compared the mean ratings for each teacher in Treatment D, while the second analysis was employed to test differences among the mean teacher ratings in Treatment E. The resulting F ration yielded in both of these analyses was significant, as reported in Tables 18 and 19. The Scheffé S-method was employed to further test which of the mean teacher ratings contributed most to the significant between-teacher F ratio present in both treatments. This technique resulted in the identification of Teacher 5 as the only teacher to differ significantly from the other teachers in mean teacher ratings on items 6-11 in either Treatment D or E. In both treatments, the mean rating of Teacher 5 on these "general teaching" items was lower than the mean rating of any other teacher. The mean rating for Teacher 5 was significantly lower than the means of three of the remaining seven teachers in Treatment D and was significantly lower than the mean ratings of all of the other seven teachers in Treatment E. Table 20 shows the K value obtained by use of the Scheffé S-method in each of these inter-teacher mean rating comparisons.

Analysis of Teaching Behavior as Measured by stR

This rating device was scaled so that an index score of 100 reflected pupil perception of maximum "high discovery-low expository" teaching behavior, while an index score of 0 reflected pupil perception of maximum "low discovery-high expository" teaching behavior.

In order to assess pupil perception of teaching behavior on the discovery-expository dimension, both before and after the experimental period, the stR was administered in both the pre- and post-test series.

The pre-test scores of necessity reflected pupils' perceptions of the teacher's typical behavior in teaching arithmetic prior to the experimental period. It was predicted that typical teaching behavior, as measured by the stR index scores, would be found to combine some elements of both discovery and expository teaching but would resemble most closely the expository teaching model. The mean stR pre-test index score for each treatment is reported in Table 9. Although the prediction that typical teaching behavior would resemble most closely the expository teaching model was borne out (pre-test means of 42.9, 44.5, and 42.2), the experimenter was surprised at how closely the pre-test means approached the theoretical mean of 50 which would reflect discovery and expository teaching in equal proportions.

The post-test index scores obtained by use of this instrument should have reflected, in some measure at least, pupils' perceptions of the teacher's typical behavior in teaching arithmetic during the experimental period. Inasmuch as the rating scale was constructed so that the "pure discovery" model would be ranked as 100 on the scale and the "pure expository" model would be ranked as 0, it was predicted that the pre- to post-test gains in teacher index scores would show the following trends: (a) a positive pre- to post-test in gain in Treatment D, (b) a negative pre- to post-test gain in Treatment E, and (c) relatively little change in the pre- to post-test scores in Treatment C. The mean stR pre- to post-test gains for each teacher in Treatments D and E are presented in Table 21 and show a definite gain for each experimental treatment group in the predicted direction.

From this pre- to post-test shift in pupil perception of the specified teaching behaviors, it was inferred that all teachers followed the prescribed teaching models relatively closely, with the exception of Teachers 3 and 8 in their teaching in Treatment E. In these two cases, the gain was in the opposite direction to that predicted. The difference in magnitude between the total treatment gains was judged to be a function of initial pupil perception of all the eight teachers as exhibiting more expository than discovery characteristics in their previous teaching of arithmetic.

The stR gain scores were analyzed, by teacher and treatment, in order to assess the significance of any differences. One analysis of variance compared the teacher gain scores in Treatment D with the teacher gain scores in Treatment E. The results of this analysis, reported in Table 22, show a highly significant difference between the gains in the two treatments, again validating the proposition that the teachers were able to vary their teaching behaviors sufficiently to effect a real test of the two teaching models.

An analysis of variance comparing teacher differences in their stR gain scores within Treatment D yielded no significant differences. The results of this analysis, shown in Table 23, lend credence to the notion that all teachers presented the Treatment D materials in a reasonably comparable manner. A similar analysis of variance comparing teacher differences in Treatment E did yeild an F ratio significant at the .1 level. The results of this analysis are presented in Table 24. The Scheffé S-method was employed to test which of the mean teacher gains contributed most to the obtained F value. The use of this technique, which is less apt to accept chance differences than use of the F statistic, revealed no significant differences between any of the mean teacher stR pre- to post-test gain scores.

The mean stR pre- to post-test gain scores for Treatments D, E, and C are shown in Table 9. Intertreatment comparisons of these means were made by use of the t statistic. The mean gain score for Treatment D was positive and differed significantly from the mean gains in Treatments E and C, both of which were negative. The difference between mean gains in Treatments E and C was not significant, although the negative shift from pre- to post-test scores was greater in Treatment E than in Treatment C. Table 25 presents the values of t obtained in each of these comparisons.

Tests of Hypotheses

Two separate analyses of covariance were used to analyze the data contributed by each of the instruments designed to test the specific hypotheses. The first analysis which is reported for each instrument was a two-way factorial design, by treatment and teacher, analyzing the data from the two experimental treatments, Treatments D and E. The second analysis was a simple analysis of covariance with treatment as the main effect. This analysis included the control group and was, therefore, not performed on the nsTT and apTT, as these tests were not given in the control group. The covariates used in each analysis are shown in the table which reports the results of that analysis.

The two-way analysis which used both treatment and teacher as main effects yielded significant first-order interactions (teacher-by-treatment) and significant differences between teachers on all nine instruments analyzed. No definitive explanation could be given for either of these results. The experimenter would speculate that variables of teaching behavior and personality too fine to be detected by the gross measures of teaching behavior used in this study were operative in all classes and that these were largely responsible for the significant findings. It would also be speculated that some teacher personalities were more compatible with one of the experimental treatments than with another and that this resulted in the teacher-by-treatment interaction effects. Further research to investigate these findings is proposed, using refinements of the measures of teaching behavior employed in this study. No further attempt at explanation of these results is given here.

One purpose of this study was to identify fruitful directions that future research might take. Therefore, no differences which were found to be significant beyond the .10 level were rejected out-of-hand as being due to chance. Rather, these findings have been reported on a tentative basis. Definite judgment concerning such differences found to be significant between the .10 and .05 levels has been withheld until such time as further research should have the opportunity of corroborating or contradicting these tenuous findings.

Test of Hypothesis 1 - Initial Learning

Hypothesis 1. There is a between-treatment difference, in favor of Treatment D, in initial learning as measured by the mean scores on a criterion test of the selected mathematical concepts given immediately after instruction.

The criterion test used to measure initial learning was the TOT-CT post-test. The mean score and standard deviation for each class in Treatments D and E, as well as the initial and adjusted total treatment means, are presented in Table 26.

A two by eight, treatment-by-teacher, analysis of covariance was performed on the TOT-CT post-test scores; this analysis is presented in Table 27. The F values obtained for both main effects and for interaction (treatment-by-teacher) were significant. The between-treatment F ratio provided a direct test of the hypothesis and was found to be significant at the .01 level. The difference was, however, in the direction opposite to that predicted by the experimenter.

A second analysis of covariance, comparing the TOT-CT post-test scores in all three treatments yielded a highly significant F value. This analysis is presented in Table 28.

The use of the t statistic was employed to test differences between the individual means. This technique resulted in two findings:

(a) both Treatments D and E differed significantly from the control group on the TOT-CT post-test scores; and (b) there was a significant differences, in favor of Treatment E, between the experimental treatments. Table 29 summarizes these intertreatment comparisons.

In light of this evidence, the hypothesis regarding initial learning was rejected. It was concluded that the expository method (Treatment E) was significantly superior to the discovery method (Treatment D) in producing results on an initial learning test.

Test of Hypothesis 2 - Retention at Two Weeks

Hypothesis 2. There is a between treatment difference, in favor of Treatment D, in retention as measured by the mean scores on a criterion test of the selected mathematical concepts given two weeks after instruction.

The criterion test used to measure retention was the TOT-RT. The mean score and standard deviation for each class in Treatments D and E, as well as the initial and adjusted total treatment means, are presented in Table 30.

A two by eight, treatment-by-teacher, analysis of covariance was performed on the TOT-RT scores; this analysis is presented in Table 31. The computed F ratios for both mean effects and for interaction (treatment-by-teacher) were significant. The between-treatment F ratio provided a direct test of the hypothesis and was found to be significant at the .05 level. The difference was in favor of Treatment D and, therefore, supported the hypothesis.

A second analysis of covariance, comparing the TOT-RT scores in all three treatments, yielded a highly significant F value. This analysis is presented in Table 32.

The use of the t statistic was employed to test differences between the individual means. This technique resulted in two findings: (a) both Treatments D and E differed significantly from the control group on the TOT-RT scores; and (b) there was a significant difference, in favor of Treatment D, between the experimental treatments. The difference between Treatments D and E reached the .02 level of significance in this comparison. Table 33 summarizes these intertreatment comparisons.

On the basis of this evidence, the hypothesis regarding retention at two weeks was accepted.

Test of Hypothesis 3 - Concept Transfer

Hypothesis 3. There is a between-treatment difference, in favor of Treatment D, in concept transfer as measured by the mean scores on a transfer test of the selected mathematical concepts given immediately after instruction.

The criterion test used to measure concept transfer learning was the cTT. The mean score and standard deviation for each class in Treatments D and E, as well as the initial and adjusted total treatment means for this instrument, are presented in Table 34.

A two-way, treatment-by-teacher, analysis of covariance was performed on the cTT scores; this analysis is presented in Table 35. The F values obtained for teacher effect and interaction were both found to be significant. The difference between Treatment D and E scores on this instrument was found to be significant at the .10 level, in favor of Treatment D.

A simple analysis of covariance comparing the cTT scores in all three treatments yielded an F value significant at the .10 level. This analysis is presented in Table 36.

Comparison of pairs of the individual means by the t test showed the following results: (a) no significant differences between Treatments C and E, (b) a significant difference at the .10 level between Treatments D and C, and (c) a significant difference at the .05 level between the mean cTT scores of pupils in Treatments D and E. These comparisons are presented in Table 37.

This evidence, although rather tenuous, favors the discovery method (Treatment D). This interpretation seems reasonable when the very small difference between Treatments E and C is considered. Hypothesis 3 was tentatively accepted until such time as further research should confirm or contradict these results.

Test of Hypothesis 4 - Negative Concept Transfer

Hypothesis 4. There is a between-treatment difference, in favor of Treatment D, in negative concept transfer as measured by the mean scores on a negative concept transfer test given immediately after instruction.

The cnTT was the criterion measure used to measure negative concept transfer. The data, contributed by this instrument are shown in Tables 38, 39, 40, and 41. There is no support in these data for the notion of a difference between treatments in negative concept transfer. Neither of the computed F ratios for treatment effects reached the values required for any acceptable level of significance, although the between teachers and interaction differences were found to be statistically significant. It is interesting to note that the control group did not differ significantly from either of the experimental groups on this instrument.

On the basis of this information, hypothesis 4 was rejected.

Test of Hypothesis 5 - Attitude

Hypothesis 5. There is a between-treatment difference, in favor of Treatment D, in attitude as measured by the mean scores of attitudinal criterion tests given immediately after instruction.

Three criterion tests were used to measure pupil attitude toward arithmetic. These were the sdAT post-test, the asAT post-test and the TOT-AT post-test. The mean score and standard deviation for each class in Treatments D and E, as well as the initial and adjusted total treatment means for each of these instruments, are presented in Tables 42, 46, and 50. The results of the teacher-by-treatment analysis of covariance for each instrument are shown in Tables 43, 47, and 51. The results of the simple analyses of covariance for all three treatments on these instruments are shown in Tables 44, 48, and 52, while the comparisons of individual treatment means for each instrument by the t technique are presented in Tables 45, 49, and 53.

In the two-way analyses of covariance, a significant first order interaction (teacher-by-treatment) was found for each of the three attitude measures. Each measure also yielded a significant F ratio for the between teachers effects. None of the differences between Treatments D and E yielded a significant F ratio, although each difference was in the predicted direction.

In the simple analyses of covariance, the comparison between the three treatment groups yielded statistically significant differences on all three instruments. Comparison of treatment means by t tests showed both experimental groups to differ significantly from the control group on all three attitude measures.

The obtained value of t in the Treatment D vs. Treatment E comparison on the TOT-AT measure was significant at the .1 level. This was only one of six attitude comparisons between Treatments D and E, however, and the remaining five did not reach a minimum acceptable level of significance.

All of these differences were in the predicted direction, however, and the mean pre- to post-test attitude gains shown by pupils in Treatment D were considerably greater than the accompanying gains shown in Treatment E. Despite these provocative findings, the data were judged to not lend sufficient support to the hypothesis to permit its acceptance. Hypothesis 5 was, therefore, rejected.

Test of Hypothesis 6 - Transfer of Heuristics

Hypothesis 6. There is a between-treatment difference, in favor of Treatment D, in the transfer of heuristics as measured by the mean scores on two heuristic transfer tests given immediately after instruction.

Two criterion tests were used to measure the effect of the two treatments on pupil ability to transfer heuristics to a new task situation. They were the nsTT and apTT.

The nsTT purported to measure pupil ability to transfer heuristics to a new paper-and-pencil task. The mean score and standard deviation for each class in Treatments D and E, as well as the initial and adjusted total treatment means for this instrument, are presented in Table 54.

A teacher-by-treatment analysis of covariance was performed on the nsTT scores; this analysis is presented in Table 55. Significant between-teacher and significant interaction (treatment-by-teacher) effects were found for this instrument. A significant difference between the two treatments reached the .05 level of significance, in favor of Treatment D.

The apTT was used to measure pupil ability to transfer heuristics to a new task presented by the teacher. The mean score and standard deviation for each class in Treatments D and E, as well as the initial and adjusted total treatment means for this instrument, are presented in Table 56.

An analysis of covariance using teachers and treatments as main effects was conducted with the apTT scores of subjects in Treatment D and E. The results of this analysis are presented in Table 57. Again, significant F ratios were obtained for between teacher effects and teacher-by-treatment interaction. The predicted difference between the means of the two treatment groups was found to be significant at the .025 level.

In light of this evidence, hypothesis 6 was accepted.

Discussion of Results

Teacher Behavior. Some of the most significant results of the study were those which indicated that the experimental teachers were able to consciously shift their teaching behavior from one teaching model to the other while still retaining a high degree of fidelity to each model. Not only did the raters agree that teacher adherence to both theoretical teaching models was high, but this judgment was corroborated by pupil ratings of teaching behavior during the experimental period. Of most importance for the interpretation of the results of this study was the clear-cut evidence that the subjects in the two experimental treatments received instruction by two consistently different methods of teaching, each of which closely paralleled the particular model prescribed.

Initial Learning and Retention. The most dramatic finding in this study was the rather startling reversal in rank of Treatments D and E between the administration of the TOT-CT post-test and the TOT-RT administered two weeks later. Although Treatment E was significantly superior to Treatment D on the test of initial learning ($p < .01$, the retention test given two weeks later showed Treatment E not only to have lost this initial superiority but also to have been surpassed by Treatment D. The pupils taught by the discovery method were able to retain significantly more material ($p < .05$) over the two-week period, notwithstanding the fact that they had evidenced knowledge of significantly less material than the Treatment E group on the test of initial learning. This finding strongly suggests that presentation of mathematical concepts to sixth-grade pupils by techniques of discovery teaching causes the learner to integrate the content into his conceptual framework in such a manner that he can retain it more readily than if the concepts had been presented to him by an expository teaching method.

Concept Transfer. Despite a pre-experimental attempt to develop an instrument that would be appropriate to measure concept transfer, it was the experimenter's opinion that the results yielded would have been more explicit had this transfer test been modified in two ways.

First, the scoring system was inadvertently devised in a manner which made it theoretically possible for a pupil to receive 30 per cent of the total score through accurate computation alone, without any knowledge of the selected mathematical principles. Pupils in the control group received a mean score on this test which represented 17 per cent of the total points possible. This result was judged to be largely due to the inept weighting of the scoring system.

Second, it was judged that this transfer test may have been too difficult for the subjects involved, thus preventing a range in achievement commensurate with ability and other variables. Some evidence of the relatively high difficulty level of this test was provided by the teachers who overheard subjects commenting to this effect following the test administration. Additional evidence was found in the mean score for subjects in both experimental treatments. This score, part of which could have been attributable to the weighting of the scoring system mentioned above, was less than 46 per cent of the maximum possible score on this test. It was the opinion of the teachers who administered and scores this test that the test difficulty led to considerable random guessing on the part of the subjects.

Based on the above evidence and inference, it is the opinion of the experimenter that the cTT was subject to considerable random error of measurement. These random errors could act to increase the analysis of covariance error term, thus making it more difficult to obtain a significant F ratio. The obtained between-treatment F ratio in the teacher-by-treatment analysis of covariance favored Treatment D over Treatment E at a minimum acceptable level of significance ($p < .10$). In a second analysis of covariance comparing all three treatments, the obtained F ratio reached the same level of significance. A t test, based on means and standard errors of the mean adjusted by this covariance technique, showed the Treatment D group to be significantly superior to the Treatment E group at the .05 level. As a result of these findings, the experimenter would speculate that modifications of the instrument to reduce the random error of measurement would have resulted in more highly significant differences in favor of the discovery method employed in this study.

Negative Concept Transfer. The hypothesis that Treatment D would be favored on a test of negative concept transfer was based on the rationale that pupils who found empirically that a principle did not apply in some situation would be more cautious in overgeneralizing the principle to inappropriate situations than pupils who had been told where the rule could not be used and who had never suffered the discomfiture of applying it inappropriately. This hypothesis was not supported by the data.

There were no differences between the mean scores of Treatments D, E, or C on the cnTT. This finding should allay concerns of proponents of both discovery and exposition, for neither treatment seemed to create rigidity or fixation of concept transfer in this study.

Transfer of Heuristics. In light of the evidence yielded by both the nsTT and the apTT, it seems reasonable to conclude that learning by discovery techniques significantly increases pupil ability to use discovery problem-solving approaches in new situations, both those which require paper-and-pencil application and those which involve verbal presentation by the teacher. Treatment D was shown to be significantly superior to Treatment E on both of these dimensions in the present study.

Attitude. The results yielded by the attitude measures were somewhat equivocal. Only one of the six comparisons between Treatments D and E reached the .10 level of significance. The other five comparisons were not statistically significant although the differences were all in the direction predicted. It was interesting to note that the pre- to post-test gains on the attitude measures also consistently favored Treatment D.

A post-experimental evaluation of the research project also yielded provocative, although subjective, results related to the above. Among other questions, the eight experimental teachers were asked which of their two classes seemed to like the "new math" better. Six of the eight teachers responded that their Treatment D group gave considerably greater expressions of liking the new arithmetic program than did their Treatment E class. The remaining two teachers indicated that both of their classes seemed to like the arithmetic content equally well.

The considerations mentioned in the preceding paragraphs, while not offered as conclusive evidence, were judged by the experimenter to be sufficiently perturbing to necessitate an attempt at explanation. Several explanations of these findings are possible:

(1) The discovery-expository dimension of teaching has no real effect upon pupil attitude toward arithmetic as measured by attitude scales of expressed likes, dislikes, etc. This explanation seems unlikely because the differences, one of which reached significance at the .10 level, and the pre- to post-test gains were all in a direction which favored Treatment D. This explanation would also conflict with the judgment of the majority of teachers who participated in this study and who were closest to actual expressions of pupil likes and dislikes of the task content.

(2) The measures of attitude used were inadequate to detect real differences in attitude. This explanation is plausible because of two factors: (a) the low six-week test-retest reliability of .44 for the sdAT, and (b) face validity being the only validity offered for these instruments. This explanation seems unlikely to completely account for the seemingly incongruous findings, however, as five of the six test-retest reliabilities on these instruments seem adequate and reasonable validity could be inferred from the careful development, by authors of the original versions, of the original measures from which these instruments were adapted.

(3) The "arithmetic" post-test series was seen by the pupils as a boring repetition of a fairly tedious "arithmetic" pre-test series, thus resulting in some "test antipathy" toward arithmetic in any form. This explanation was substantiated by subjective evidence of pupil comments overheard by several of the teachers during and after the post-test administration. While the pre-test administration of the criterion measures was likely novel, it was probable that the subjects found

little intrinsic interest in the post-test series. Several teachers observed that pupils were more careless in following instructions on the post-test series and finished the tests much more quickly than in the pre-test administration. Although the attitude measures were administered at the beginning of the post-test series, the pupils were told just prior to administration of the first attitude test that they would receive the entire test battery again plus four new arithmetic tests. This statement was followed by immediate negative reactions from many pupils. To the extent that these reactions would cause a decrease in motivation, one result would liekly be an increase in random errors due to carelessness. Another result would seem to be negative shift in expressed attitude toward arithmetic because of hostility toward the tedious second administration of these tests. This notion is partially supported by the observation that the control groupshowed a substantial negative pre- to post-test shift on all three attitude measures. This general phenomenon, though temporary, could have operated to mask real differences in attitude between the experimental groups to a point where the majority of comparisons failed to yield statistically significant differences.

(4) The marked change from arithmetic textual materials used prior to the experiment to the instructional materials used in this study tended to cause a definite improvement of pupil attitude toward arithmetic because of a shift in pupil perception of the subject matter content. There is evidence to support this explanation. Both experimental treatment groups showed substantial increases in mean attitude scores on all three measures. This may be due to the explanation given above, or to operation of the so-called "Hawthorne Effect" in both experimental treatments. Whatever the cause, such a notable increase in attitude in

both treatments very probably reflected some constant uncontrolled variable much more than the specified treatment variables. To the extent that this assumption is correct, real differences between the treatments would be at least partially masked.

(5) The six-week instructional period was not long enough to effect real changes in pupil attitude toward arithmetic. Some arguments can be presented both for and against this explanation. The significant differences between the control group and the two experimental groups on all three attitude measures might tend to suggest that a six-week instructional period was sufficient to effect change in pupil attitude. However, to the extent that negative testing effects proposed in (3) operated in all treatment groups, but were counterbalanced in the experimental groups by factors proposed in (4), this could also explain the significant differences between the control group and the experimental groups on the attitude tests.

The present study cannot provide evidence as to which of these, or other, explanations is best. It is the opinion of the experimenter, however, that an explanation in terms of (3) and (4) is superior to an explanation in terms of (1), (2), or (5).

Number of Units Completed. It should be noted that the varying rate of presentation among teachers resulted in some variation among schools in the amount of the sequenced instructional materials completed during the experimental period. There was virtually no variation between treatments, however, when summed across schools. In seven of the eight schools, the number of instructional units completed by the two treatment groups were equal. The single exception was in the school where fifth-grade pupils were used as subjects. In this school, the Treatment E class completed two pages more of the instructional materials than the Treatment D group in the same period of time. This difference was discounted by the experimenter as inconsequential.

The variation between schools in the number of units completed could also be a highly plausible explanation of the significant between-teacher F ratio yielded by all measures used in this study and should be added to other possible explanations of this finding given in the section on "Tests of Hypothesis" earlier in this chapter.

TABLE 1

CHARACTERISTICS OF SAMPLE BY SCHOOL AND TREATMENT: GRADE, SEX, AND NUMBER

School	Treatment	Grade	Sex		N
			Male	Female	
1	D	6	14	13	27
	E	6	11	16	27
2	D	5	14	13	27
	E	5	9	18	27
3	D	6	13	14	27
	E	6	13	14	27
4	D	6	12	15	27
	E	6	16	11	27
5	D	6	14	13	27
	E	6	13	14	27
6	D	6	12	15	27
	E	6	13	14	27
7	D	6	17	10	27
	E	6	11	16	27
8	D	6	14	13	27
	E	6	14	13	27
9	C	6	60	46	106
Schools 1-8	D		110	106	216
	E		100	116	216
	D & E		210	222	432
All Schools D, E, & C			270	268	538

TABLE 2

CHARACTERISTICS OF SAMPLE BY SCHOOL AND TREATMENT: IQ, ARITHMETIC COMPUTATION, AND ARITHMETIC PROBLEM-SOLVING MEANS AND STANDARD DEVIATIONS

School and Treatment		Arithmetic Computation		Arithmetic Problem-solving		IQ	
		M	SD	M	SD	M	SD
1	D	7.56	.86	8.94	1.06	119.00	11.34
	E	7.24	.70	8.58	1.40	120.41	13.25
2	D	4.55[a]	.88	4.55[a]	.88	101.56	14.46
	E	4.87[a]	.88	4.87[a]	.88	102.30	12.61
3	D	5.88	.84	5.72	1.02	96.11	12.36
	E	5.54	.99	5.51	1.16	93.56	12.96
4	D	6.51	.89	7.37	1.48	113.07	15.96
	E	6.45	.92	6.91	1.46	111.44	15.54
5	D	5.93	.51	5.69	.79	99.82	8.81
	E	5.53	.78	5.52	1.19	99.59	12.91
6	D	6.07	.89	6.31	1.50	100.04	14.56
	E	6.09	1.07	6.18	1.62	102.00	21.59
7	D	6.26	1.21	6.23	1.50	103.37	16.88
	E	6.20	1.17	6.00	1.47	97.78	14.49
8	D	6.95	.73	8.01	1.22	113.19	12.59
	E	6.74	.72	7.78	1.25	113.12	16.79
9	C	5.87	.72	6.38	1.31	102.25	13.62
1-8	D	6.21	1.19	6.60	1.78	105.77	15.42
	E	6.08	1.15	6.42	1.75	105.03	17.26
	D & E	6.15	1.17	6.51	1.77	105.40	16.35

[a] The arithmetic computation and arithmetic problem-solving scores in School 2 were based on norms for grade 5.

TABLE 3

FINAL PAGE PRESENTED TO EACH TREATMENT GROUP*

School	Treatment D	Treatment E
1	A1.67	A2.72
2	A1.54	A2.63
3	A1.51	A2.58
4	A1.67	A2.72
5	A1.51	A2.58
6	A1.59	A2.65
7	A1.67	A2.72
8	A1.67	A2.72

*Pagination refers to Appendices related to this study but presented in Part II of the final report published as a separate volume.

TABLE 4

RELIABILITY FOR EACH INSTRUMENT[a]

Instrument	Reliability: Test-retest Consec. Days	Test-retest Six-week	Parallel Form Consec. Days	Parallel Form Six-week	Other
cbAT	.91	.13			
sdAT	.87	.44[d]			
asAT	.78	.75			
TOT-AT[b]	.84	.68			
stR	.92	.92			
TOT-CT	.70[c]	.73[c]			
TOT-RT	.75[c]		.75[c]	.69[c]	
cTT	.82				
cnTT	.64[d]				
nsTT					See text discussion under "Instruments"
apTT					

[a] Per cent of rater agreement for olR, taR is reported in Table 9.

[b] TOT-AT is the summation of sdAT and asAT.

[c] This was judged to be a very acceptable reliability inasmuch as the range of scores was from 0 to 6 on the consecutive day test-retest and 0 to 22 on the six-week test-retest.

[d] These discouragingly low reliability estimates were judged to be less than acceptable, even for making group comparisons. Partial cause of this low reliability was presumed to be due to the relatively small number of items in these instruments.

TABLE 5

CORRELATION MATRIX FOR CONCEPT, RETENTION, AND CONCEPT TRANSFER MEASURES WITH IQ AND ARITHMETIC ACHIEVEMENT MEASURES

Measures	1	2	3	4	5	6
1. TOT-CT		.76	.67	.62	.67	.67
2. TOT-RT			.72	.62	.66	.69
3. cTT				.63	.69	.75
4. IQ					.67	.74
5. Arithmetic Computation						.88
6. Arithmetic Problem Solving						

Note: These values were computed with the data from Treatments D and E combined.

TABLE 6

CORRELATION MATRIX FOR HEURISTIC TRANSFER TESTS WITH IQ AND ARITHMETIC ACHIEVEMENT MEASURES

Measures	1	2	3	4	5
1. nsTT		.50	.49	.48	.55
2. apTT			.62	.64	.72
3. IQ				.67	.74
4. Arithmetic Computation					.88
5. Arithmetic Problem Solving					

Note: These values were computed with the data from Treatments D and E combined.

TABLE 7

PER CENT OF RATER AGREEMENT ON olR, taR

Between	Situation			Mean Between Raters Agreement
	Live Rating	Live vs. Tape Rating	Tape Rating	
Raters 1 and 2	91 82 91 91		82	87.4
Raters 1 and 3		55 64	64	61.0
Raters 2 and 3		73	82	77.5
Mean Situation Agreement	88.8	64.0	76.0	Mean % of Rater Agreement 75.8

TABLE 8

NUMBER OF PUPILS EXCLUDED AND NUMBER OF CASES ADDED: BY SCHOOL AND TREATMENT

School	Treatment	Original N	Randomly Excluded	Cases Added at Mean	Resulting N
	D	30	3		27
	E	30	3		27
	D	33	6		27
	E	34	7		27
	D	23		4	27
	E	24		3	27
	D	30	3		27
	E	29	2		27
	D	21		6	27
	E	24		3	27
	D	28	1		27
	E	30	3		27
	D	29	2		27
	E	30	3		27
	D	29	2		27
	E	27		27	
Totals	D	223	17	10	216
	E	228	18	6	216
	D & E	451	35	16	432

TABLE 9

INITIAL PRE- AND POST-TEST MEANS AND STANDARD DEVIATIONS: BY TREATMENT

Measure	Treatment D		Treatment E		Treatment C	
	$\overline{X}$	SD	$\overline{X}$	SD	$\overline{X}$	SD
TOT-CT pre-test	4.1	4.8	3.9	5.1	4.8	3.7
TOT-CT post-test	52.5	17.1	54.0	18.0	5.0	5.2
sdAT - pre-test	14.3	7.1	15.9	8.6	15.3	7.1
sdAT - post-test	16.3	7.9	16.8	8.4	14.6	8.9
asAT pre-test	23.2	19.5	28.8	21.8	31.4	18.7
asAT post-test	30.0	21.6	32.2	20.1	28.8	23.1
TOT-AT pre-test[b]	37.6	24.6	44.9	28.9	46.7	23.9
TOT-AT	46.3	27.9	49.1	26.7	43.9	29.1
cnTT	10.8	6.2	10.6	6.0	10.3	5.5
cTT	34.0	19.5	31.5	18.3	12.7	6.4
nsTT	21.1	11.4	18.8	10.7	c	c
apTT	18.5	11.5	15.9	11.6	c	c
TOT-RT	48.9	20.9	46.7	21.3	6.7	6.5
stR pre-test	42.9	18.6	44.5	18.3	42.2	21.2
stR gain	7.4	2.8	-1.6	2.9	-.7	.1
olR, taR 1-5	4.6	.4	1.4	.4	d	d
olR, taR 6-11	4.6	.4	4.5	.3	d	d
Arith. Comp.	6.2	1.2	6.1	1.1	5.9	.7
Arith. Prob. Solv.	6.6	1.8	6.4	1.8	6.4	1.3
IQ	105.8	15.4	105.0	17.3	102.2	13.6

[a] Prior to adjustment by the covariance technique.
[b] TOT-AT is the summation of asAT and sdAT.
[c] Not administered to Treatment C. See page 5.31.
[d] Treatment C teacher not rated.

TABLE 10

COMPARABILITY OF TREATMENT GROUPS ON IQ, ARITHMETIC COMPUTATION, ARITHMETIC PROBLEM SOLVING, AND PRETEST MEASURES

Measure	Obtained Values of t		
	D with E	D with C	E with C
IQ	.47	2.51*	1.87
Arith. Comp.	1.16	3.58***	2.28*
Arith. Prob. Solv.	1.18	1.47	.27
TOT-CT Pre-test	.398	1.650	2.026*
sdAT Pre-test	2.108*	1.473	.791
asAT Pre-test	2.814**	4.457***	1.590
TOT-AT Pre-test	2.829**	3.906***	.703
stR Pre-test	.301	.122	.401

*Significant at the .05 level for a two-tailed test.

**Significant at the .01 level for a two-tailed test.

***Significant at the .001 level for a two-tailed test.

TABLE 11

MEAN TEACHER RATINGS ON ITEMS 1-5 OF olR, taR SCALE: BY TEACHER AND TREATMENT

Treatment	Teacher								Total Treatment
	1	2	3	4	5	6	7	8	
D	4.76	4.68	4.73	4.32	4.38	4.53	4.73	4.45	4.56
E	1.40	1.32	1.20	1.56	1.30	1.56	1.20	1.68	1.41

TABLE 12

ANALYSIS OF VARIANCE OF MEAN TEACHER RATINGS ON ITEMS 1-5 OF olR, taR: BETWEEN TREATMENTS D & E

Source of Variation	Sum of Squares	df	Mean Square	F	p
Treatments	180.4	1	180.4	1061.18	<.001
Within	12.0	71	.17		
Total	192.4	72			

TABLE 13

ANALYSIS OF VARIANCE OF DIFFERENCES BETWEEN MEAN TEACHER RATINGS IN TREATMENTS D & E ON ITEMS 1-5 OF olR, taR AND MAXIMUM RATINGS FOR EACH TEACHING MODEL ON THESE ITEMS

Source of Variation	Sum of Squares	df	Mean Square	F	p
Treatments	.1	1	.10	.59	n.s.
Within	12.0	71	.17		
Total	12.1	72			

TABLE 14

ANALYSIS OF VARIANCE OF MEAN TEACHER RATINGS IN TREATMENT D ON ITEMS 1-5 OF olR, taR: BETWEEN TEACHERS

Source of Variation	Sum of Squares	df	Mean Square	F	p
Teachers	.9	7	.13	1.00	n.s.
Within	3.4	26	.13		
Total	4.3	33			

TABLE 15

ANALYSIS OF VARIANCE OF MEAN TEACHER RATINGS IN TREATMENT E ON ITEMS 1-5 OF olR, taR: BETWEEN TEACHERS

Source of Variation	Sum of Squares	df	Mean Square	F	p
Teachers	1.3	7	.19	.90	n.s.
Within	6.4	31	.21		
Total	7.7	38			

TABLE 16

MEAN TEACHER RATINGS ON ITEMS 6-11 OF olR, taR SCALE: BY TEACHER AND TREATMENT

Treatment	1	2	3	4	5	6	7	8	Total Treatment
D	4.58	4.57	4.77	4.40	4.08	4.70	4.82	4.65	4.57
E	4.58	4.47	4.60	4.60	3.75	4.50	4.75	4.68	4.49

TABLE 17

ANALYSIS OF VARIANCE OF MEAN TEACHER RATINGS ON ITEMS 6-11 OF olR, taR: BETWEEN TREATMENTS D & E

Source of Variation	Sum of Squares	df df	Mean Square	F	p
Treatments	.2	1	.20	1.43	n.s.
Within	12.8	94	.14		
Total	13.0	95			

TABLE 18

ANALYSIS OF VARIANCE OF MEAN TEACHER RATINGS IN TREATMENT D ON ITEMS 6-11 OF olR, taR: BETWEEN TEACHERS

Source of Variation	Sum of Squares	df	Mean Square	F	p
Teachers	2.4	7	.34	4.88	$<.01$
Within	2.7	40	.07		
Total	5.1	47			

TABLE 19

ANALYSIS OF VARIANCE OF MEAN TEACHER RATINGS IN TREATMENT E ON ITEMS 6-11 OF olR, taR: BETWEEN TEACHERS

Source of Variation	Sum of Squares	df	Mean Square	F	p
Teachers	4.1	7	.59	6.56	<.01
Within	3.6	40	.09		
Total	7.7	47			

TABLE 20

OBTAINED K VALUE IN EACH COMPARISON OF TEACHER 5 MEAN RATING ON ITEMS 6-11 WITH MEAN RATING OF OTHER TEACHERS ON ITEMS 6-11 OF olR, taR: BY TREATMENT

Teacher	Treatment	1	2	3	4	6	7	8
5	D	3.29	3.22	4.54*	2.11	4.08*	4.87**	3.75
5	E	4.83**	4.19*	4.94**	4.94**	4.36*	5.81***	5.41**

*Significant at the .05 level.

**Significant at the .01 level.

***Significant at the .001 level.

TABLE 21

MEAN PRE- TO POST-TEST GAINS ON stR: BY TEACHER AND TREATMENT

Treatment	1	2	3	4	5	6	7	8	Total Treatment
D	8.1	7.0	12.8	6.6	4.7	7.1	3.2	9.6	7.39
E	-1.4	-.48	1.6	-1.7	- .4	-4.8	-4.5	3.5	-1.56

TABLE 22

ANALYSIS OF VARIANCE OF TEACHER stR PRE- TO POST-TEST GAIN SCORES: BETWEEN TREATMENTS D AND E

Source of Variation	Sum of Squares	df	Mean Square	F	p
Treatments	3393.1	1	3393.1	25.589	$<.001$
Within	52772.6	398	132.6		
Total	56165.7	399			

TABLE 23

ANALYSIS OF VARIANCE OF TEACHER stR PRE- TO POST-TEST GAIN SCORES IN TREATMENT D: BETWEEN TEACHERS

Source of Variation	Sum of Squares	df	Mean Square	F	p
Teachers	1518.8	7	217.0	1.478	n.s.
Within	28184.6	192	146.8		
Total	29703.4	199			

TABLE 24

ANALYSIS OF VARIANCE OF TEACHER stR PRE- TO POST-TEST GAIN SCORES IN TREATMENT E: BETWEEN TEACHERS

Source of Variation	Sum of Squares	df	Mean Square	F	p
Teachers	1653.5	7	236.2	2.118	$<.1$
Within	21415.8	192	115.5		
Total	23069.3	199			

TABLE 25

COMPARISON OF TREATMENT GROUPS ON MEAN stR PRE- TO POST-TEST GAIN SCORES

Measure	Obtained Values of t		
	D with E	D with C	E with C
stR	5.93*	7.80**	.77

*Significant at .001 level for a one-tailed test with 14 df.
**Significant at .001 level for a one-tailed test with 9 df.

TABLE 26

MEANS AND STANDARD DEVIATIONS OF TOT-CT POST-TEST SCORES OF SUBJECTS IN TREATMENTS D AND E: BY TEACHER AND TREATMENT

Treatment		1	2	3	4	5	6	7	8	Total Init.	Treatment Adj.
D	$\overline{X}$	70.2	46.7	48.1	53.8	40.5	32.8	59.5	68.7	52.54	51.51
	SD	4.3	15.3	13.3	16.9	11.0	10.5	11.3	9.8	17.14	
E	$\overline{X}$	68.9	45.2	44.7	55.6	42.2	42.1	63.3	70.2	54.03	54.16
	SD	6.5	14.8	18.5	15.4	13.3	20.5	12.0	17.99		

TABLE 27

ANALYSIS OF COVARIANCE OF TOT-CT POST-TEST SCORES OF SUBJECTS IN TREATMENTS D AND E BY TEACHER AND TREATMENT

Source of Variation	Sum of Squares	df	Mean Square	F	p
Treatment	721.54	1	721.54	7.435	<.01
Teacher	23444.48	7	3349.21	34.510	<.001
Interaction	1427.79	7	203.97	2.102	<.05
Within	39983.31	412	97.05		
Total	65577.12	427			

Note: IQ, Arith. Comp. Arith, Prob. Solving, and TOT-CT pre-test scores used as covariates.

TABLE 28

ANALYSIS OF COVARIANCE OF TOT-CT POST-TEST SCORES OF SUBJECTS IN TREATMENTS D, E, AND C

Source of Variation	Sum of Squares	df	Mean Square	F	p
Treatment	174395.99	2	87198.00	669.63	<.001
Within	69145.41	531	130.22		
Total	243541.40	533			

Note: IQ, Arith. Comp., Arith. Prob. Solving, and TOT-CT pre-test scores used as covariates.

TABLE 29

COMPARISON OF MEAN TOT-CT POST-TEST SCORES, ADJUSTED BY COVARIANCE, FOR TREATMENTS D, E, AND C

Treatment	Initial Mean	Adjusted Mean	SE of Adj. Mean
D	52.54	51.51	0.779
E	54.03	54.16	0.778
C	4.97	6.82	1.125

Measure	Obtained values of *t*		
	D with E	D with C	E with C
TOT-CT	2.41*	32.62**	34.55**

*Significant at the .01 level for a one-tailed test.
**Significant at the .001 level for a one-tailed test.

TABLE 30

MEANS AND STANDARD DEVIATIONS OF TOT-RT SCORES OF SUBJECTS IN TREATMENTS D AND E: BY TEACHER AND TREATMENT

Treatment		Teacher								Total	Treatment
		1	2	3	4	5	6	7	8	Init.	Adj.
D	$\bar{X}$	70.7	46.4	48.3	36.9	29.3	41.1	48.2	69.8	48.88	42.66
	SD	3.6	19.9	16.6	16.7	12.6	21.7	20.3	5.4	20.88	
E	$\bar{X}$	69.9	41.7	40.3	46.4	30.7	36.4	42.3	65.9	46.71	40.13
	SD	5.2	19.2	17.8	20.6	13.6	23.9	19.5	8.7	21.30	

TABLE 31

ANALYSIS OF COVARIANCE OF TOT-RT SCORES OF SUBJECTS IN TREATMENTS D AND E: BY TEACHER AND TREATMENT

Source of Variation	Sum of Squares	df	Mean Square	F	p
Treatment	451.99	1	451.99	3.918	<.05
Teacher	16944.28	7	2420.61	20.983	<.001
Interaction	3802.32	7	543.19	4.709	<.001
Within	47529.33	412	115.36		
Total	68727.92	427			

Note: IQ, Arith. Comp., Arith, Prob. Solv, and TOT-CT post-test scores used as covariates.

TABLE 32

ANALYSIS OF COVARIANCE OF TOT-RT SCORES OF SUBJECTS IN TREATMENTS D, E, AND C

Source of Variation	Sum of Squares	df	Mean Square	F	p
Treatment	2642.68	2	1321.34	9.85	<.001
Within	71235.71	531	134.15		
Total	73878.39	533			

Note: IQ, Arith. Comp., Arith. Prob. Solving, and TOT-CT post-test scores used as covariates.

TABLE 33

COMPARISON OF MEAN TOT-RT SCORES, ADJUSTED BY COVARIANCE, FOR TREATMENTS D, E, AND C

Treatment	Initial Mean	Adjusted Mean	SE of Adj. Mean
D	48.88	42.66	0.860
E	46.71	40.13	0.912
C	6.68	32.78	1.981

Measure	Obtained Values of t		
	D with E	D with C	E with C
TOT-RT	2.02*	4.57**	3.37**

*Significant at the .02 level for a one-tailed test.
**Significant at the .001 level for a one-tailed test.

TABLE 34

MEANS AND STANDARD DEVIATIONS OF cTT SCORES OF SUBJECTS IN TREATMENTS D AND E: BY TEACHER AND TREATMENT

Treatment		Teacher								Total Treatment	
		1	2	3	4	5	6	7	8	Init.	Adj.
D	$\overline{X}$	53.3	19.9	21.9	31.1	25.1	33.3	34.3	53.3	34.02	30.28
	SD	11.4	15.7	12.9	16.8	11.6	18.0	19.8	15.6	19.50	
E	$\overline{X}$	52.5	25.5	19.1	29.0	23.9	19.5	35.9	46.6	31.49	28.16
	SD	13.4	13.3	11.2	15.2	14.0	16.3	15.9	14.4	18.27	

TABLE 35

ANALYSIS OF COVARIANCE OF cTT SCORES OF SUBJECTS IN TREATMENTS D AND E: BY TEACHER AND TREATMENT

Source of Variation	Sum of Squares	df	Mean Square	F	p
Treatment	360.19	1	360.19	3.089	<.10
Teacher	5573.51	7	796.22	6.827	<.001
Interaction	3703.95	7	529.14	4.537	<.001
Within	48046.62	412	116.62		
Total	57584.26	427			

Note: IQ, Arith. Comp., Arith. Prob. Solving, and TOT-CT post-test scores used as covariates.

TABLE 36

ANALYSIS OF COVARIANCE OF cTT SCORES OF SUBJECTS IN TREATMENTS D, E, AND C

Source of Variation	Sum of Squares	df	Mean Square	F	p
Treatment	628.07	2	314.04	2.603	<.10
Within	64063.68	531	120.65		
Total	64691.75	533			

Note: IQ, Arith. Comp., Arith. Prob. Solving, and TOT-CT scores used as covariates.

TABLE 37

COMPARISON OF MEAN cTT SCORES, ADJUSTED BY COVARIANCE, FOR TREATMENTS D, E, AND C

Treatment	Initial Mean	Adjusted Mean	SE of Adj. Mean
D	34.02	30.28	0.816
E	31.49	28.16	0.865
C	12.73	27.12	1.879

	Obtained values of t		
Measure	D with E	D with C	E with C
cTT	1.78**	1.54*	.50

*Significant at the .10 level for a one-tailed test.
**Significant at the .05 level for a one-tailed test.

TABLE 38

MEANS AND STANDARD DEVIATIONS OF cnTT SCORES OF SUBJECTS IN TREATMENTS D AND E: BY TEACHER AND TREATMENT

		Teacher								Total Treatment	
Treatment		1	2	3	4	5	6	7	8	Init.	Adj.
D	$\bar{X}$	12.0	8.9	9.6	10.9	8.7	11.5	9.8	14.6	10.76	10.58
	SD	5.8	6.3	5.5	6.2	3.7	7.8	5.8	6.0	6.16	
E	$\bar{X}$	11.3	11.1	9.3	11.6	11.7	6.7	10.2	13.1	10.63	10.67
	SD	4.5	5.3	6.3	6.5	4.7	7.1	5.8	6.1	6.02	

TABLE 39

ANALYSIS OF COVARIANCE OF cnTT SCORES OF SUBJECTS IN TREATMENTS D AND E: BY TEACHER AND TREATMENT

Source of Variation	Sum of Squares	df	Mean Square	F	p
Treatment	3.03	1	3.03	.098	n.s.
Teacher	600.09	7	85.73	2.779	<.01
Interaction	542.52	7	77.50	2.512	<.025
Within	12741.30	413	30.85		
Total	13886.94	428			

Note: IQ, Arith. Comp., and Arith. Prob. Solving scores used as covariates.

TABLE 40

ANALYSIS OF COVARIANCE OF cnTT SCORES OF SUBJECTS IN TREATMENTS D, E, AND C

Source of Variation	Sum of Squares	df	Mean Square	F	p
Treatment	1.31	2	0.65	.02	n.s.
Within	16731.38	532	31.45		
Total	16732.69	534			

Note: IQ, Arith. Comp., and Arith. Prob. Solving scores used as covariates.

TABLE 41

MEAN cnTT SCORES, ADJUSTED BY COVARIANCE, FOR TREATMENT D, E, AND C

Treatment	Initial Mean	Adjusted Mean	SE of Adj. Mean
D	10.76	10.58	0.383
E	10.63	10.67	0.382
C	10.28	10.55	0.551

TABLE 42

MEANS AND STANDARD DEVIATIONS OF sdAT POST-TEST SCORES OF SUBJECTS IN TREATMENTS D AND E: BY TEACHER AND TREATMENT

		Teacher								Total Treatment	
Treatment		1	2	3	4	5	6	7	8	Init.	Adj.
D	$\overline{X}$	15.6	15.5	19.3	15.7	16.8	17.0	9.1	21.3	16.29	16.70
	SD	6.4	8.6	4.8	5.4	6.4	6.7	12.2	5.4	7.95	
E	$\overline{X}$	22.7	19.1	17.2	14.5	13.6	17.4	9.2	20.9	16.83	16.43
	SD	3.7	5.8	8.1	7.8	10.3	8.3	8.7	5.1	8.43	

TABLE 43

ANALYSIS OF COVARIANCE OF sdAT POST TEST SCORES OF SUBJECTS IN TREATMENTS D AND E: BY TEACHER AND TREATMENT

Source of Variation	Sum of Squares	df	Mean Square	F	p
Treatment	5.95	1	5.95	.161	n.s.
Teacher	1779.87	7	254.27	6.863	<.001
Interaction	494.93	7	70.70	1.908	<.1
Within	15263.62	412	37.05		
Total	17544.44	427			

Note: IQ, Arith. Comp., Arith. Prob. Solving, and sdAT pre-test scores as covariates.

TABLE 44

ANALYSIS OF COVARIANCE OF sdAT POST-TEST SCORES OF SUBJECTS IN TREATMENTS D, E, AND C

Source of Variation	Sum of Squares	df	Mean Square	F	p
Treatment	324.86	2	162.43	3.58	<.05
Within	24090.80	531	45.37		
Total	24415.66	533			

Note: IQ, Arith. Comp., Arith. Prob. Solving, and sdAT pre-test scores used as covariates.

TABLE 45

COMPARISON OF MEAN sdAT POST-TEST SCORES, ADJUSTED BY COVARIANCE, FOR TREATMENTS D, E, AND C

Treatment	Initial Mean	Adjusted Mean	SE of Adj. Mean
D	16.29	16.70	0.461
E	16.83	16.43	0.460
C	14.60	14.60	0.662

Measure	Obtained values of t		
	D with E	D with C	E with C
sdAT post-test	.41	2.60**	2.27*

*Significant at the .02 level for a one-tailed test.
**Significant at the .001 level for a one-tailed test.

TABLE 46

MEANS AND STANDARD DEVIATIONS OF asAT POST-TEST SCORES OF SUBJECTS IN TREATMENTS D AND E: BY TEACHER AND TREATMENT

Treatment		Teacher 1	2	3	4	5	6	7	8	Total Treatment Init.	Adj.
D	$\overline{X}$	31.1	35.6	39.2	29.4	29.1	28.1	11.4	35.9	29.98	32.67
	SD	18.7	24.7	10.7	19.2	18.4	19.7	30.3	15.9	21.61	
E	$\overline{X}$	45.5	30.1	39.3	29.8	25.2	32.5	14.7	40.9	32.25	31.05
	SD	11.4	23.9	18.1	20.0	21.8	16.9	17.6	11.9	20.09	

TABLE 47

ANALYSIS OF COVARIANCE OF asAT POST-TEST SCORES OF SUBJECTS IN TREATMENTS D AND E: BY TEACHER AND TREATMENT

Source of Variation	Sum of Squares	df	Mean Square	F	p
Treatment	217.48	1	217.48	1.173	n.s.
Teacher	4028.52	7	575.50	3.105	<.01
Interaction	3320.85	7	474.41	2.560	<.025
Within	76366.14	412	185.35		
Total	83932.99	427			

Note: IQ, Arith. Comp., Arith. Prob. Solving, and asAT pre-test scores used as covariates.

TABLE 48

ANALYSIS OF COVARIANCE OF asAT POST-TEST SCORES OF SUBJECTS IN TREATMENTS D, E, AND C

Source of Variation	Sum of Squares	df	Mean Square	f	p
Treatment	3225.65	2	1612.82	7.79	<.001
Within	109887.51	531	206.95		
Total	113113.16	533			

Note: IQ, Arith. Comp., Arith. Prob. Solving, and asAT pre-test scores used as covariates.

TABLE 49

COMPARISON OF MEAN asAT POST-TEST SCORES, ADJUSTED BY COVARIANCE, FOR TREATMENTS D, E, AND C

Treatment	Initial Mean	Adjusted Mean	SE of Adj. Mean
D	29.98	32.67	0.992
E	32.25	31.05	0.982
C	28.85	25.80	1.424

Measure	Obtained values of *t*		
	D with E	D with C	E with C
asAT post-test	1.17	3.95*	3.03*

*Significant at the .001 level for a one-tailed test.

TABLE 50

MEANS AND STANDARD DEVIATIONS OF TOT-AT POST-TEST SCORES OF SUBJECTS IN TREATMENTS D AND E: BY TEACHER AND TREATMENT

		Teacher								Total Treatment	
Treatment		1	2	3	4	5	6	7	8	Init.	Adj.
D	$\overline{X}$	46.7	51.1	58.4	45.1	45.9	45.2	20.6	57.2	46.28	49.61
	SD	23.5	31.7	14.1	23.5	23.6	23.9	40.4	19.8	27.92	
E	$\overline{X}$	68.2	49.2	56.5	44.3	38.9	49.9	23.9	61.8	49.08	47.26
	SD	13.6	27.7	23.4	26.3	30.8	23.7	24.6	14.6	26.74	